THE UNIVERSITY
of EDINBURGH

Re-engineering *E. coli* for *in vivo* production of fluorometabolites

Konstantinos Markakis

THE UNIVERSITY *of* EDINBURGH

Thesis submitted for the degree of

Doctor of Philosophy

University of Edinburgh

School of Engineering

Year of Submission 2019

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Konstantinos Markakis

24 July 2019

Abstract

Fluorinated products are widely used in pharmaceutical and agrochemical applications where a large portion of the commercial compounds being developed contain at least one atom of fluorine. Most of the conventional fluorination reactions require volatile chemicals and are highly exothermic requiring controlled procedures which are rather costly, produce low yields and exhibit low selectivity of the moiety accepting fluorine. As a result, much attention has been drawn on enzymatic fluorination in a quest to ameliorate these issues.

The only characterized enzyme that mediates direct fluorination is the 5'-fluoro-5'-deoxyadenosine (5'-FDA) synthase (a.k.a. fluorinase), which uptakes an *S*-adenosyl-L-methionine (SAM) molecule and a fluorine atom as substrates and yields 5'-FDA and methionine. Variants of this enzyme have been extensively studied structurally and functionally with *in vitro* assays and they exhibit very high homology, and a similarly slow activity.

In this work, a bacterial *in vivo* bioreactor is presented, based on a variant of the enzyme, which is found in *Streptomyces* sp. MA37 (FlA1). The base host that has been chosen for the implementation of modifications is *Escherichia coli* BL21(DE3), which is a host of choice for the over-expression of recombinant proteins and in some occasions utilized for the construction of novelty bioreactors.

In the case of enzymatic fluorination, a few challenges had to be addressed.

First of all, *E. coli* contains an ion channel protein, termed CrcB, which expels fluoride anions out of the cell. Secondly, SAM cannot passively enter *E. coli* cells rendering the intracellular molarity of SAM a bottleneck. Last but not least, the *E. coli* purine nucleoside phosphorylase, DeoD, has been suggested to degrade the fluorination product, 5'-FDA. Therefore, the modified *E. coli* strain, consists of 2 deleted genes ($\Delta crcB \Delta deoD$), the expression of a SAM transporter from *Rickettsia prowazekii* and the over-expression of *flA*1. Using this strain, *in vivo* production of 5'-FDA to near millimolar concentrations has been confirmed using an array of chemical analysis methods including High Performance Liquid Chromatography (HPLC), Fluorine Nuclear Magnetic Resonance (F-NMR) and High Resolution Mass Spectrometry (HRMS).

Acknowledgements

First of all I would like to thank God for providing with the strength and tranquillity needed to complete this strenuous task. I would like to thank my supervisor Prof. Alistair Elfick, for his trust on critical decisions regarding the project and for providing me with a friendly working environment which was paramount to the success of this work.

I would like to thank co-author and collaborator Dr Aitor de las Heras for the knowledge and experience he passed on to me throughout the project. Many thanks goes to Dr Vlastimil Srsen for the invaluable support and help which always exceeded his prescribed duties as lab manager.

Special thanks to Dimitrios Tsikritsis for his continuing guidance and feedback on procedural/scientific matters and for our endless brainstorming sessions.

I would like to thank Aleksandros Kleidaras to whom I owe a great debt of gratitude for his support in difficult situations and in particular for his help in completing and submitting this thesis.

A big thank you to all those people that helped me and I somehow forgot to mention them here, their good deeds will be remembered.

Finally, many thanks to my family and friends back home, for their ongoing support for which I am - and always will be - grateful.

Abbreviations

^{19}F-NMR Fluorine-19 Nuclear Magnetic Resonance. 149, 151, 153, 154, 156–159

^{1}H-NMR Proton nuclear magnetic resonance. 149, 151, 154

GP Gaussian Processes. 5, 6, 44, 45, 47–50, 52, 60, 61, 69, 123, 127–129, 133, 136, 143, 146, 183

4-FT 4-fluorothreonine. 5, 17, 18, 20–22, 29, 183, 185

5'-ClDA 5'-chloro-5'-deoxyadenosine. 6, 7, 25, 161, 179, 181

5'-FDA 5'-fluoro-5'-deoxyadenosine. v, vi, 6, 7, 17, 18, 22, 25, 106–108, 124, 125, 142, 149–151, 153, 154, 156–159, 161, 179–182, 184, 185

5'-FDI 5'-fluoro-5-deoxyinosine. 18, 153, 154

5-FDR 5-fluoro-5-deoxy-ribose. 20, 27

5-FDRP 5-fluoro-5-deoxy-D-ribose 1-phosphate. 18, 27, 106, 153

5-FDRibulP 5-fluoro-5-deoxy-D-ribulose 1-phosphate. 18

5-FDRul 5-fluoro-5-deoxy-ribulose. 20

5-FHPA (2*R*3*S*4*S*)-5-fluoro-2,3,4-trihydroxypentanoic acid. 26, 27

ANOVA Analysis Of Variance. 123, 129, 175, 176, 178, 179

CARS Coherent Anti-Stokes Raman Spectroscopy. 42, 179, 185

CSV comma-separated values. 60

DCW Dry Cell Weight. 158

DE Directed Evolution. 4, 5, 13, 14, 124, 179

DSB Double Strand Break. 32, 33, 102–104, 109

FAc Fluoroacetate. 15–17, 21, 27, 30

FACS Fluorescence Activated Cell Sorting. 15

FP Fluorescent Protein. 44, 45, 55–57, 59, 62, 65, 69, 70, 74, 183, 184

FU fluorescence intensity units (FU), in arbitrary units. 53, 54, 60, 65, 67

FWHM Full-Width Half-Maximum. 167, 168

GFP Green Fluorescent Protein. 54, 65, 69

GUI Graphical User Interface. 5

HPLC High Performance Liquid Chromatography. 149, 151, 154

HRMS High Resolution Mass Spectrometry. 149, 151, 154

KO knockout. 91, 94, 104, 109–111, 139, 144, 145

LB Lysogeny Broth. 63, 65, 67, 81–85, 100, 102–104, 116, 118, 119, 127, 130, 134, 148–151, 177, 178

MAT methionine adenolyltranferase. 111, 112, 114

MTAP *S*-methyl-5'-thioadenosine phosphorylase. 18, 28

OD Optical Density. 5, 44, 46–48, 50, 52–57, 59, 60, 63, 65, 67, 69, 74, 75, 84, 85, 127–130, 133, 134, 137, 139, 141–143, 146, 148, 150, 158

ORF Open Reading Frame. 18, 27, 28, 106

PCR Polymerase Chain Reaction. 14, 34, 38–40, 44, 63, 78, 82, 87, 88, 96–98, 100, 102, 103, 109, 111, 117, 119, 149

PEP phosphoenolpyruvate. 89

PNP Purine Nucleoside Phosphorylase. 18, 78, 105–108, 124, 137, 149, 153, 154, 156, 159

PTS phosphotransferase system. 89

RBS Ribosome Binding Site. 64, 65, 69, 78, 87, 116, 120, 184

RFP Red Fluorescent Protein. 87

RS Raman Spectroscopy. 4–6, 41, 42, 161, 172, 179, 185

S/N supernatant. 37, 150, 151, 159, 177

SAH *S*-adenosyl-L-homocysteine. 24, 114–116

SAM *S*-adenosyl-L-methionine. v, vi, 6, 7, 17, 22–25, 27, 28, 78, 111–118, 120, 124, 134, 144–146, 148, 150, 151, 154, 156, 158–161, 172–182, 184–186

SB Synthetic Biology. 9, 10, 12, 14, 31, 43

SERS Surface Enhanced Raman Spectroscopy. 185

SRS Stimulated Raman Spectroscopy. 42, 179, 180, 185

YFP Yellow Fluorescent Protein. 63, 67

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Organofluorines and applications

Fluorinated chemistry refers to the chemistry of compounds that contain at least one atom of fluorine. Specifically, organic fluorinated chemistry refers to the compounds with an organic/carbon backbone. The unique characteristics of the carbon-fluorine bond render it very enticing for a number of applications. One characteristic is the high thermal and metabolic stability of this chemical bond resulting from its high energy, the highest among all carbon based bonds. Other attributes include low width of the bond, low electronegativity etc. which contribute to its thermal stability and hydrophobicity. There are numerous fluorinated compounds that have been synthesized and constitute products or part of them in everyday life such as refrigerants, fluorosurfactants, liquid crystal displays (LCDs), metered dose propellants. However, the most prevalent applications occur in biochemistry, agrochemistry and medicinal chemistry.

The application of fluorinated carbons is very extensive in agrochemistry with a percentage of 30-40% of newly developed agrochemicals comprising fluorinated

compounds [11]. The wide use of fluorinated compounds in this field includes insecticides, pesticides, herbicides and acaricides. A variety of popular fluorinated products that have been developed are fluorinated heterocyclic compounds. Depending on the particular pest that an agrochemical treatment is targeting, there are several fluorinated carbon alternatives categorized per functionality. Examples of herbicides include fluorinated compounds that inhibit catalysis of lipid sythesis that are essential for cell membrane construction. A category of fungicides includes compounds that inhibit mitochondrial respiration. Chitin synthase inhibitors constitute a representative example of insecticides, where potent fluorinated chemicals halt chitin synthesis in insects, a key element of their exoskeleton [12]. The main functional theme of all these modified fluorinated bioproducts is the inhibitory action of each particular chemical which acts in a key biochemical pathway of an unwanted organism or cellular process, disrupting it and ultimately halting its reproduction. This inhibition is made possible from the extreme hydrophobicity of the fluorine - carbon bond leading to lipophilicity which enhances its bioavailability.

In most cases, the inhibition process involves binding of a fluorinated analogue in an enzyme' s active site. Under normal circumstances, this binding involves a natural substrate with which the reaction proceeds normally. Substituting one or more atoms in suitable positions of the natural substrate with fluorine, or other fluorinated subgroups, results in the binding of this unnatural substrate in the active site which cannot be processed by the enzyme leading to inhibition. This general principle is similar in medical products. After the first fluorine containing drug, fludrocortisone [13] was approved back on 1954 as an anti-inflammatory agent, fluorinated compounds have been extensively introduced into the approved drugs list. Up to 2010, almost 1/3 of the 100 top-selling drugs incorporated fluorine and up to 1/5 of the approved drugs contained at least one molecule of fluorine [14]. Recently, many biological drugs, such as antibodies, have entered the top selling diminishing the contribution of fluorine drugs as top sellers, but

the overall percentage of newly approved fluorine containing drugs, exhibits an increasing trend reaching up to 30% [15]. A uracil analog, namely 5-Fluoro-Uracil (5-FU), has been extensively utilized as a chemotherapeutic agent for the treatment of several forms of cancer. In this case, with the introduction of a fluorine atom in the 5-position of uracil, the latter is transformed to an anti-metabolite which stops proliferation of cancer cells by inhibiting certain key enzymes and in turn causing metabolic imbalances that result in DNA damage [16]. Fluoxetine is a widely used anti-depressant which was first approved on 1987 and its mechanism of action, termed selective serotonin re-uptake inhibition (SSRI), is tied to selective inhibition of a small molecule efflux channel. The intricacies of the mechanism are still poorly understood, however the resulting effect is an increase of the extracellular pool of serotonin, lack of which leads to neural dysfunction. Generally, the main theme of fluorinated organic molecules is based on the anti-metabolic function of fluorine-substituted carbon-hydrogen bonds of otherwise essential molecules for the function that needs to be blocked or indirectly promoted.

Most of the conventional procedures involving halogenation often require volatile chemicals and conditions that are expensive to achieve. In particular, fluorine mediated reactions are highly exothermic and require controlled procedures which are rather costly and produce low yields [17]. Another drawback of traditional fluorination methods is the low selectivity of the carbon and the bond that will uptake the fluoride. Especially when fluorination involves a complex molecule, which is usually the case when the average complexity of target metabolites is taken into account, low selectivity becomes a serious issue as the fluorine atom can substitute a larger set of bonds. The purification procedures for the desired product are added to the list of hurdles. In an effort to tackle with the above issues, much attention has been recently drawn into enzymatic fluorination. In this case, the reaction is mediated by a particular category of enzymes known as fluorinases. Although fluoride is abundant in nature, the natural mechanisms evolved for its

processing are limited and so are these enzymes, with 5 variants having been characterized to date, all of them with low performance for large-scale industrial use. The increasing number of genomes being sequenced will undoubtedly yield more fluorinases (see section 2.3.3), however this does not negate the fact that they occur rarely. The increasing trend in demand of fluorinated products does not allow mere dependence on nature's generosity for the desired fluorinase variants, therefore, a solution must be searched within the realm of engineered enzymes. Directed Evolution (DE) is an established method for the creation of improved variants of an enzyme, or variants with altered functionality (see section 2.2 for details).

1.2 Objectives

The main purpose of this project is to explore the potential for *in vivo* fluorination by utilizing the *E. coli* bacterium, enabling its subsequent use either as a host for DE or for engineering fluorination pathways. While the *in vivo* participation of the fluorinase has been shown in another actinomycete with compatible metabolism (see section 2.3.4), there are currently no studies indicating the transfer of this functionality in an evolutionary distant species such as *E. coli*. Furthermore, as demonstrated in a recent review [18], there is no precedent for direct fluorination in *E. coli*, though the utility of such a metabolic feat has been previously highlighted [19].

E. coli constitutes a model organism for many types of studies because it has been extensively studied and well-documented, therefore, the construction of *in vivo* fluorination can also serve as a convenient genotype-phenotype coupling for DE studies. The potential of genetic selection using the constructed *E. coli* will also be explored. Alternatively, Raman Spectroscopy (RS) will be tested for suitability as a screening method.

1.3 Chapter overview

The material and outcomes of this project are organized as follows:

Chapter two consists of a description from a theoretical perspective of the tools that are extensively utilized in this work and a critical review of the literature related to biological fluorination. First of all, a brief description of synthetic biology, chassis, the concept of orthogonality and the importance of regulatory elements is given. An overview of the workflow adopted by DE experiments follows. Then the review of biological fluorination with a focus on the developments of the 4-fluorothreonine (4-FT) pathway is presented, also mentioning the indirect fluorination pathways implemented in *E. coli*. Next, three methods of gene editing are described for *E. coli* discussing their pros and cons. A summary of everyday molecular biology techniques for lab works is given afterwards. Finally, the basics of RS are outlined.

Chapter three is based on the material published by the author of this work, on the paper "Analytical approach for the calculation of promoter activities from fluorescent protein expression data" [20]. The introduction highlights qualitative and quantitative characteristics of promoters and other regulatory elements, providing reasoning why the metric used from most studies is inaccurate. The application and improvement of a robust method for growth curve fitting based on Gaussian Processes (GP) is shown followed by the development of a metric which captures the time-dependent promoter activity based on curve fitting of Optical Density (OD) and fluorescence time-series data, also with GP. A Graphical User Interface (GUI) tool incorporating these methods is presented and validation of the new metric with experimental data from different fluorescent proteins is given. A discussion on other interesting characteristics of bacterial growth curves and a summary of outcomes and future work finalize this chapter.

Chapter four contains the rationale and description of modifications that are needed for the realization of *in vivo* direct fluorination with recombinant fluorinase expression in *E. coli*. Construction of the plasmid accommodating the fluorinase gene and confirmation of high levels of expression are illustrated. Focus is then moved to the fluoride resistance mechanism of *E. coli*, a selective anion efflux channel, followed by a step-by-step report on its deletion. The same is done for the purine nucleoside phosphorylase encoded by the *deoD* gene. Last but not least, a case is made for the increase of the intracellular pool of SAM, which is facilitated by expression of a SAM transporter from *R. prowazekii*. A catalogue of strains and plasmids used or created in this chapter resides in the end.

Chapter five deals with the effects that fluorinase expression exerts on *E. coli* BL21(DE3) by fitting experimental growth curves with *GP* as laid out in chapter 3. Specifically, emphasis is given on the quite unusual finding that overexpression of the fluorinase elicits a positive effect on growth which not only persists even in the *crcB* deletion strain, but it appears to counterbalance the toxicity of intracellular fluoride. The effect of increasing concentration of extracellular SAM on strains expressing the SAM transporter is also illustrated and the tolerance of strains in high levels of SAM is confirmed. The achievement of a major milestone in this project is then presented, *in vivo* fluorination in *E. coli*. The necessity of both *crcB* deletion and SAM transporter expression is validated through a series of confirmatory experiments.

Chapter six explores the potential of RS for the detection of intracellular metabolites. The Raman spectra of key fluorination substrates and products, SAM, 5'-FDA and 5'-chloro-5'-deoxyadenosine (5'-ClDA) are reported. The ability to monitor SAM instability with RS is also illustrated. Spectral analysis of *E. coli* bacteria expressing the SAM transporter and grown in the presence of SAM in high molarities, show the potential to monitor such intracellular metabolites

with this microscopy method. Finally, the spectral differences seen in SAM, 5'-FDA and 5'-ClDA reveal a characteristic attribute which can be used for future endeavours in developing a screening method for intracellular production of 5'-FDA.

Chapter seven concludes the thesis by highlighting the prospects that open up for future endeavours based on the observations and findings of each chapter.

Chapter 2

Background and Methods

2.1 Synthetic biology and tools

Biological systems constitute the next level of material organization. Chemical reactions are made possible in biology-friendly temperatures and conditions in which critical reactions for maintenance and replication of these systems, would not happen. The catalysts for these reactions are chemical compounds themselves (enzymes) that have acquired a higher level of complexity by arranging chains of simpler molecules (amino acids) in conformations which create favourable conditions for the reactions to occur. These elaborate polymers are produced within a boundary-defined contraption (cell) in which a wide-range of interdependent control mechanisms that have evolved over the course of 4 billion years of evolution. The set of elements controlling the timing and quantity of protein expression demonstrate immense complexity and stochasticity, yet they remain incredibly robust. Synthetic Biology (SB) is a field which aspires to capture and be able to recreate this robustness in newly engineered behaviours. The engineering approach is paramount to SB and a multitude of ways for achieving this are being

continuously explored. Systems biology, is another field enforced with understanding and mechanistically modelling biological systems. Although, a flow of knowledge towards SB exists, there is a distinct difference in that the latter tries to recreate or modify behaviours which systems biology is devoted to studying. SB, as an engineering discipline should ideally try to adopt a bottom-up approach with a predefined process for building a system from its constituent parts. While there have been efforts to create a biological entity, such as a cell, totally from scratch, leading to some protocell reminiscent structures [21], the utility of these attempts is still limited. Most of SB is therefore realised in already existing natural hosts, also referred to as chassis, in harmony with the engineering view. However, the potentially unwanted interaction with the native procedures of a host, constitutes a hurdle for the predictability of an engineering approach. Therefore an important property in SB is orthogonality. An orthogonal process within a cell is a process which does not influence the cell's metabolism. This is achieved either by trial and error or by choosing simpler hosts or hosts that have been stripped of non-essential genes.

2.1.1 Bacterial Chassis

While there have been many SB studies performed with the use of mammalian or other eukaryotic hosts, with *Saccharomyces cerevisiae* being the most used from that category, the concept of orthogonality laid out above is served better with simpler (i.e. bacterial) chassis. Therefore, the better part of SB is done in bacterial hosts. Undoubtedly, the flagship of SB is *E. coli*, a natural choice in most of SB endeavours, as it is the most well-documented, extensively studied and modified bacterium, a fact which contributes to its predictability which is a desirable attribute for SB. Moreover, *E. coli* is dominantly used for everyday molecular biology tasks such as plasmid maintenance, cloning, protein extraction. In the present study, the choice to use *E. coli* BL21 (DE3) for the implementation

of modifications was made mostly because of the existing publications which describe fluorinase overexpression in this host. A number of attributes render the BL21(DE3) genotype enticing for implementation of heterologous expression, hence engineered *in vivo* activity. First of all, the T7 RNA polymerase along with the strong inducible promoter lacUV5, constitute an orthogonal expression system within the cell, which recognises foreign promoter sequences and is able to produce high quantities of a desired recombinant protein without competing for native polymerases. Secondly, the Lon and OmpT proteases are deleted, reducing protein degradation. Third, even the expression of detrimental proteins can be achieved by the strong control on expression levels from the inducible promoter.

2.1.2 Regulatory elements

The ability of biological systems to maintain their integrity and propagate in a range of environments is a defining characteristic of life. This is accomplished by a set of sophisticated control elements which determine the spatio-temporal organization of biomatter within a living system based on environmental stimuli. These regulatory elements play a key role in the effort to construct synthetic systems as they constitute the equivalent of the control circuit in a device. To further the engineering approach, an effort to standardize modular elements has yielded the Registry of Standard Biological Parts [22], a database where modular elementary parts can be submitted. Novel parts are required to follow the BioBrick^{TM} standard, a protocol which allows their assembly into more complex devices. Following the above standards, a description language for the assembly of parts *in silico* is also being developed, namely SBOL [23].

Regulatory elements are categorized in *cis* and *trans* which are terms defining the controlling element's locality in the DNA sequence compared to the controlled element. Thus, *cis*-regulatory elements are those that control a nearby sequence,

while *trans* interacts with distant sequences. Of most interest for SB are the first and are divided into the following categories:

- Promoters: They are DNA sequences which serve as recognition sites for the RNA polymerase to attach and initiate transcription of a gene or cluster of genes (operon). Different consensus sequences appear across domains. Promoters in eukaryotes are complex and they can include elements in a varying length upstream of the a gene cluster. Bacterial promoters, which are more frequently used in SB, consist of 2 consensus sequences (based on nucleotides with the highest probability of appearing) of 6 nucleotides each that lie 35 to 10 bps before the first nucleotide of the transcript. These sequences are of key importance to the strength of each promoter, for example the synthetic promoters J23100 and J23114 from the Registry of Biological Parts differ only in 2 bps in the -35 and 1 bp in the -10 position, yet they exhibit a 10-fold difference in final fluorescent output.

- Enhancers: These are DNA sequences that with the help of the appropriate transcription factor (activator) act favourably on the rate of transcription.

- Silencers: Likewise, these regions are bound by another type of transcription factor (repressor) and reduce transcription rates.

2.2 Directed evolution

The main principle governing natural evolution is simple. Out of a population of biological entities capable of replicating with random perturbations in their phenotype, individuals with higher probability of surviving in their environment will continue to replicate successfully - with higher probability - and at some point they will become dominant in the population. The controlled application of the

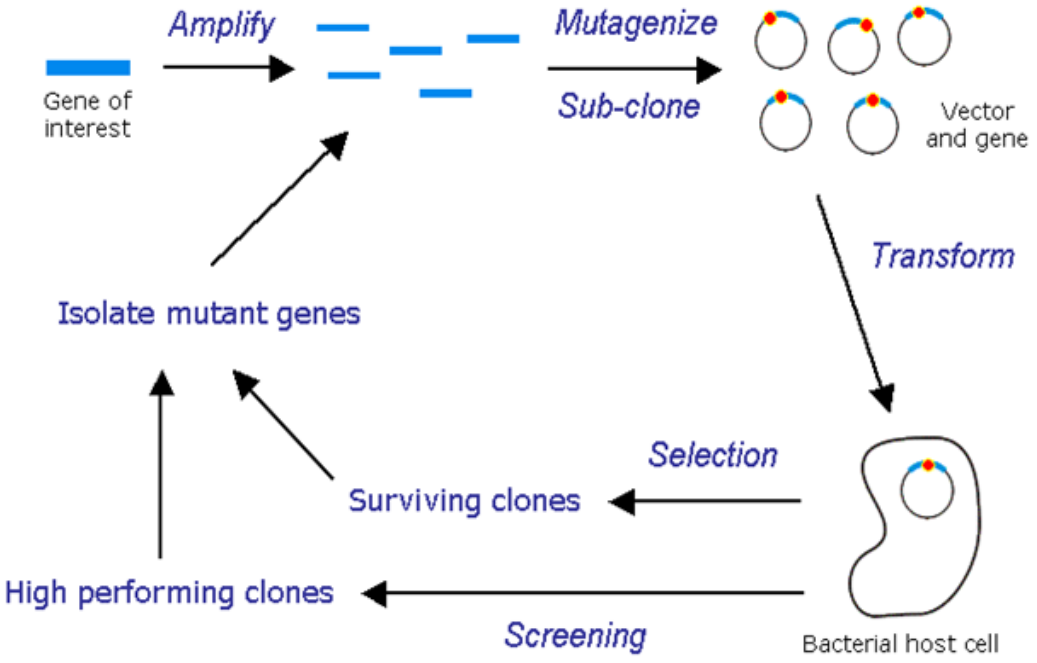


Figure 2.1: The basic workflow of DE [2].

above process in order to achieve a specific phenotypic outcome is called "Directed Evolution" (DE). The objective of the process can be either an improvement to an existing property or an alteration of the function altogether. The set of possible sequences of a particular number of amino acids is often referred to as the "sequence space", which is an N-dimensional discrete and finite space with 20 possible values in each dimension, where N is the number of amino acids in the chain. Inside the sequence space there is a sequence that exhibits the optimal capacity for a particular objective. Starting with a gene which encodes a wild type protein of interest, the main workflow of DE is shown in figure 2.1. The first step is the introduction variability in the nucleotide sequence which translates to variability in the amino acid sequence. The variants obtained from the above process comprise a mutant DNA library for that particular gene. The next step is the expression of that DNA library and screening of the resulting proteins for a particular activity of interest. The protein sequences that show an adequate

level of activity are selected and prepared for another round which starts from the initial step of the creation of the mutant library as above, expression, screening and so on.

There are 3 distinct ways to ensure genotype-phenotype coupling. One approach is to utilize the natural host of the protein of interest and use its own life cycle for selection. Another approach is the transformation of hosts that have been extensively studied and contain the necessary toolkits for high levels of expression, for example particular strains of *E. coli*. A third approach, is the *in vitro* DE of proteins. Essentially, this approach consists of an *in vitro* translation system which is coupled to the cycle of DE instead of living cells, with the amino acid chain coupled with the corresponding mRNA during the selection phase for genotype-phenotype linkage [2]. The type and extent of induced variability during the first step depends on the type of the protein, the objective of optimization or alteration and the domain of knowledge around that particular kind of protein. The most common method is error-prone Polymerase Chain Reaction (PCR) (epPCR) where the PCR occurs with suboptimal conditions introducing mutations to random positions on the amino acid chain. Another method is recombination or "DNA shuffling", a method borrowed from natural evolution, where there is a number of parent genes cut in particular positions and then reassembled in a random manner. Targeted methods include site directed mutagenesis, where mutagenic primers are designed to import mutations to specific positions in the gene (low number of mutations) and *de novo* gene synthesis, backed up by information acquired from SB approaches (high number of mutations).

Once the diversification phase is complete, selection of the variants must take place in order to pick improved variants for a particular objective. This is achieved with a selection system or screening technique. A robust, high throughput and error-free method is based on genetic selection. For example, in a competition

system where the absence of a molecule will lead to a growth defect, selection will take place directly inside the environment containing different clones. Other screening methods include Fluorescence Activated Cell Sorting (FACS), a high-throughput method, albeit limited from the necessity of fluorescent proteins, however, coupling a fluorescent reporter with the activity of a non-fluorescent substrate can broaden its use.

2.3 Biological Fluorination

Despite the fact that fluorine is the most abundant halogen found on earth, it does not occur analogously as a component of biochemically produced organic molecules. Several reasons have been proposed such as the insolubility of minerals containing fluorine or the inability of haloperoxidases, the key halogen-metabolising enzymes encountered in nature, to utilize fluoride as a halogen [8]. This is due to the unique properties of fluoride, having the highest electronegativity and the highest heat of hydration amongst all other halogens, along with a very low abundance in the oceans, limiting its bioavailability [24]. Another possible reason lies in the evolutionary route towards fluoride resistance mechanisms most of which involve fluoride expelling channels (see 4.2), further reducing its contact with cells. The rarity of biologically produced fluorinated organic molecules has driven research towards identifying these unique pathways and if possible utilizing biotechnology in order to expand the toolbox of available fluorination methods. Until the early 2000s, the exact fluorination mechanism had yet to be elucidated in spite of the first study to demonstrate the occurrence of biological fluorination dating back to 40s. The latter illustrated the existence of Fluoroacetate (FAc) in the leaves of *Dichapetalum cymosum* [25]. Later in the 60s a few other instances were highlighted in other plants yielding FAc and a few derivatives of it [26] [27], [28], [29]. The ability of some plants to

elaborate fluorometabolites is variable and in some cases even detrimental [28] indicating a non-specific mechanism of catalysis. In other species however this ability seem to have further evolved exhibiting high amounts of accumulated fluorinated compounds even in the presence of low concentrations of inorganic fluoride, for example the *Dichapetalum Braunii* demonstrates more a 100-fold higher concentration of fluoroacetic acid in the leaves of young plants than the inorganic fluoride concentration which it is exposed to [30]. The exact mechanism of fluorometabolite accumulation in plants even in these interesting cases has not been determined yet [31]. The main reason is the difficulty in obtaining plant material and in the variability and unpredictability of observed fluorination in specific plant-derived tissue cultures [32]. It has been proposed that the unpredictability and gradual loss of ability in synthesizing fluorometabolites from plant tissue cultures is the result of separation from the true source of the precursor FAc which is produced by soil bacteria [33] a handful of which are now known to produce fluorometabolites, with FAc being the prevalent end-product by a well characterized pathway with its first step mediated by the only enzyme known to perform direct fluorination of a biomolecule using inorganic fluoride (see next section). Other instances of fluorometabolites include the *Streptomyces calvus* broad-specificity antibiotic nucleocidin [34], [35]. The inherent variability followed by low titres and seemingly spontaneous halt of fluorometabolite production is also evident in this case. A recent study has shown however that production of the antibiotic is restored with the correction of a defect in Leu-tRNA(UUA) of the submitted *S. calvus* strain from the initial study [36]. However, it was reported that an *S. calvus* variant with a functional tRNA(UUA) still suffered from low titres. The radioisotope incorporations observed during incubations of this variant, though establishing some metabolic constraints [37], are inconclusive as to the specifics of the putative fluorination pathway and cannot dismiss the possibility of an effect related to the broad-specificity of a particular enzyme of a different main function.

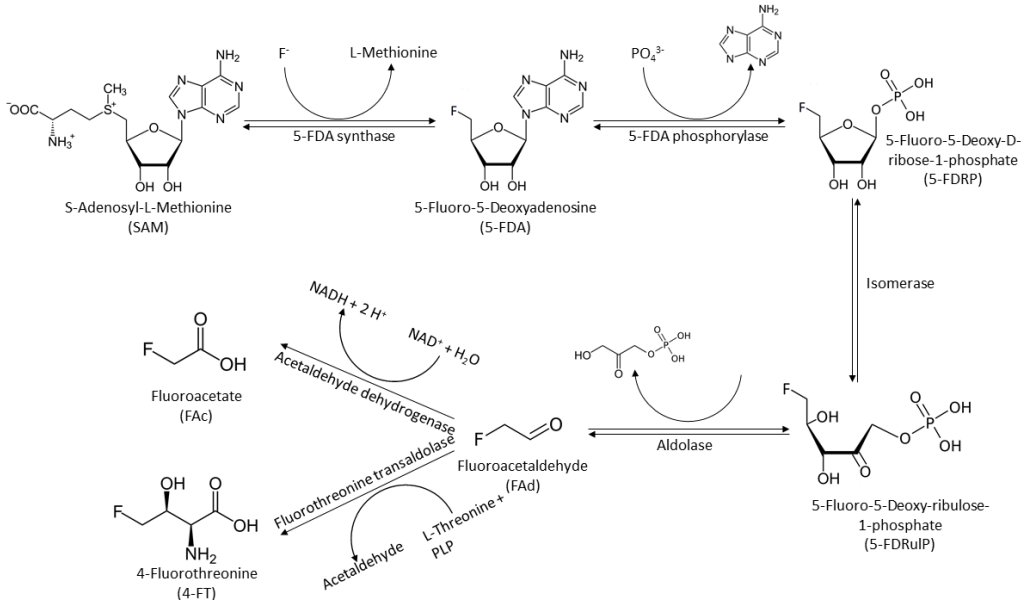


Figure 2.2: The 4-FT - Fac pathway from *S. cattleya*.

2.3.1 Fluorometabolites from bacteria: The 4-FT pathway

The first fully-characterized pathway for the production of fluorometabolites is the *Streptomyces cattleya* 4-FT - FAc pathway. This actinomycete contains a cluster of 6 enzymes which process inorganic fluoride and SAM yielding FAc and 4-FT. *S. cattleya* had initially demonstrated the ability to synthesize FAc and 4-FT in the presence of inorganic fluoride a few decades back [38], [8] but the characterization and assignment of the pathway enzymes was accomplished only during the past decade. In figure 2.2 the *S. cattleya* pathway is shown. The pathway itself has been reconstituted *in vitro* with some replacement homologues from other organisms which did not appear to hinder the pathway's capacity to elaborate 4-FT [39]. The first step of fluorinated biosynthesis was found to be a unique enzyme specifically evolved to process inorganic fluoride, termed 5'-FDA synthase or fluorinase [40], [41], and comprises the vessel for realization of direct fluorination in *E. coli* presented in the present work.

The second step in the pathway is performed by a Purine Nucleoside Phosphorylase (PNP) -type enzyme which cleaves the purine base from the molecule and replaces it with a phosphate group yielding 5-fluoro-5-deoxy-D-ribose 1-phosphate (5-FDRP). Although *S. cattleya* PNP acts on the 5'-FDA directly to yield 5-FDRP, the *E. coli* PNP (deoD-type PNP EC:2.4.2.1) reacts with 5'-fluoro-5-deoxyinosine (5'-FDI) [42], a shunt product of the pathway which is not observed *in vivo*, but is found during the cell-free extract mediated activity of the pathway. This difference is attributed to some type of compartmentalization [43]. Interestingly, the PNP from *S. cattleya* bears higher homology with the enzymes annotated as *S*-methyl-5'-thioadenosine phosphorylase (MTAP) type rather than the usual PNPs, with an identity of 38% with the human MTAP and a 35% of homology with the yeast homologue. On the contrary, *E. coli* PNP genes (xapA and deoD) exhibit 23% and 16% of homology respectively and there is no homologue in *E. coli* annotated as MTAP.

The third enzyme of the pathway, namely 5-FDRP isomerase, is an interesting case as it bears significant homology to 5-methylthio-ribose-1-phosphate isomerase, which is part of the L-methionine salvage pathway. In fact, the version of this enzyme found in *Bacillus subtilus* was found to bear around 35% homology with several putative Open Reading Frame (ORF)s from *Streptomyces coelicolor* and *Streptomyces avermitilis*.

Incubations of the protein corresponding to *S. coelicolor* isomerase homologue with 5-FDRP, were found to elaborate 5-fluoro-5-deoxy-D-ribulose 1-phosphate (5-FDRibulP), product of the third step of the 4-FT pathway. Since it has been established that the isomerase from *S. coelicolor* is able to catalyse the third step, using genome mining the authors of [39] managed to identify an ORF with 75% homology. This ORF' s product was confirmed to be the 5-FDRP isomerase of the third step of the fluorination pathway.

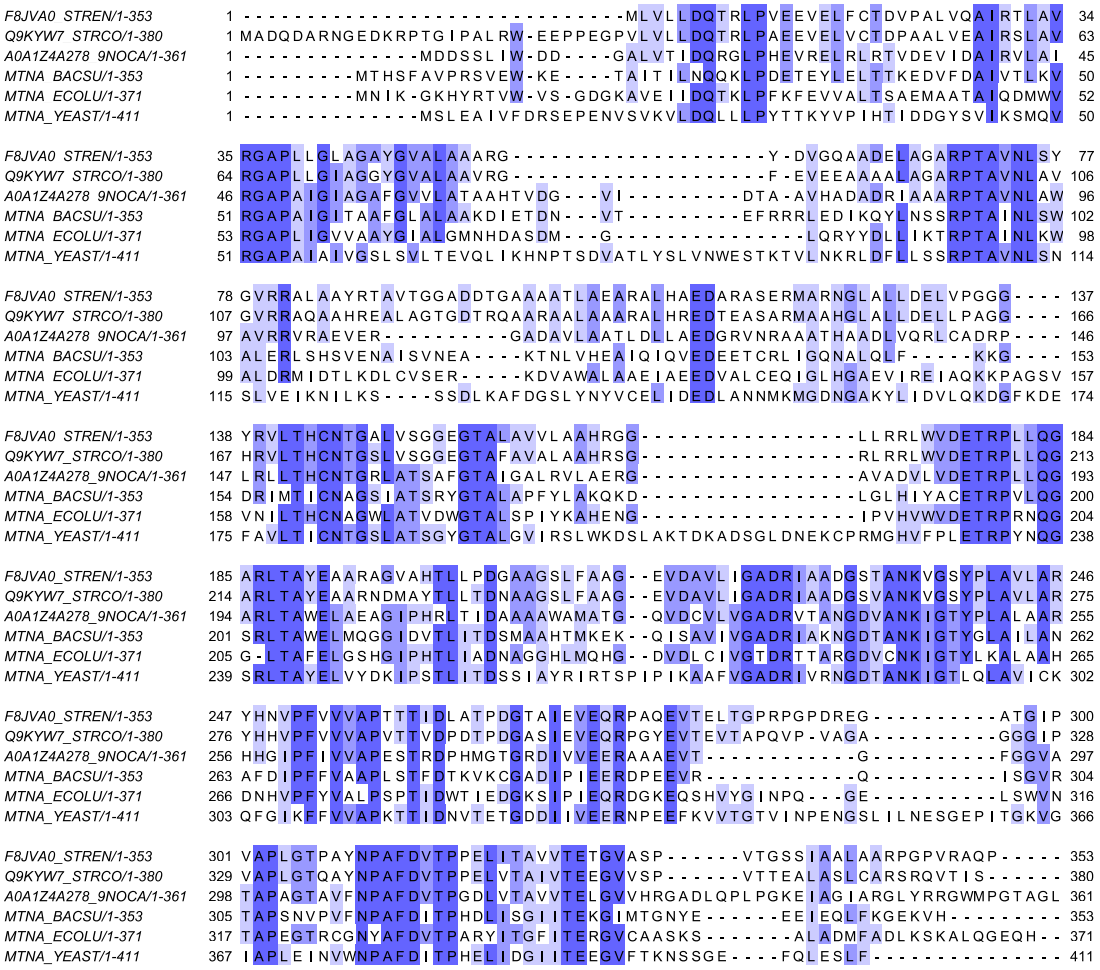


Figure 2.3: Alignment of 5-methylthio-ribose-1-phosphate isomerases from *S. cattleya*, *S. coelicolor*, *N. brasiliensis*, *B. subtilis*, *E. coli* and *S. cerevisiae*. All of them show at least 30% sequence identity with clearly conserved motifs.

This series of sequence-homology extrapolation is encouraging for the exploration of migration of fluorometabolite synthesis to other organisms, since apart from *B. subtilus* homology with the *S. cattleya* isomerase, homologues of 5-methylthio-ribose-1-phosphate isomerase exist in other biotechnologically relevant organisms such as yeast (≈33%) and *E. coli* K-12 (≈50%) (see figure 2.3) with a roughly 60% coverage in all cases. Moreover, the enzyme's function is identical as suggested by the fact that a similar glucose isomerase can efficiently catalyse the reaction 5-fluoro-5-deoxy-ribose (5-FDR) to 5-fluoro-5-deoxy-ribulose (5-FDRul) regardless of fluoride substitution in the molecule [44].

The fourth step of the pathway is performed by an aldolase. This enzyme has been reported to be elusive in its determination for the primary 4-FT host *S. cattleya* therefore a replacement fuculose aldolase from *S. coelicolor* was used and efficiently performed this fourth step of the pathway *in vitro*. Alignment of the functional fuculose aldolase reveals 3 similar genes in *S. cattleya* one of them exhibiting significant similarity (≈65%) while the other two are less similar (≈35%). Again, an *E. coli* fuculose aldolase protein was included in the alignment showing somewhat high similarity (≈40%) and the potential compatibility of the enzyme with pathway rewiring or alternative functionality in this commonly used bacterium (figure 2.4).

Finally, the last step consists of two enzymes and a crossroad for the pathway. One enzyme is the fluoroacetaldehyde dehydrogenase (*SCAT_0945*), a NAD^+ dependent enzyme which shows high affinity for fluoroacetaldehyde but also accepts glycoaldehyde and aldehyde with lower specificity. Interestingly, the yeast aldehyde dehydrogenase was also able to catalyse the reaction albeit with 4-fold less specificity [45]. Interestingly, *E. coli* accommodates a homologue for *S. cattleya* fluoroacetaldehyde dehydrogenase, namely *aldB* bearing 67% of sequence identity, much higher than that seen between the tested yeast homologue which shows a less than 40% of homology. Given the broad specificity of all these variant

dehydrogenases it is highly possible that this native dehydrogenase of *E. coli* can also catalyse this reaction. The fact that AldB can catalyse the conversion of chloroacetaldehyde even more efficiently than acetaldehyde (higher V_{max}) [46] also encouraging towards this goal. On the other hand, as also suggested in [39], the fluorothrenonine transaldolase is a unique enzyme in its own right and there is no sequence with significant similarity in *E. coli* or yeast.

Both products of the 4-FT pathway are toxic to bacterial cells. In particular, FAc is a strong inhibitor of the essential citric acid cycle. Therefore, apart from the pathway enzymes which mediate catalysis of the reactions required to yield the end-products, FAc and 4-FT, other enzymes that are critical for *S. cattleya* in order to withstand the toxicity of these fluorinated molecules are also present and their function has been elucidated. One example is the fluoroacetyl-CoA thioesterase (FlK) which mitigates FAc toxicity [47]. Until very recently, the exact inhibitory mechanism of 4-FT had yet to be illustrated. In pioneering work from Chang et al., it was assumed that 4-FT' s close resemblance to threonine could lead to misincorporations of the fluorinated version during protein synthesis. By employing bioinformatics analysis, two conserved genes have been discovered lying within the fluorination pathway locus, *fthB* and *fthC*, which code for a *trans*-acting tRNA editing protein and a 4-FT exporter respectively. Indeed, deletion of either those genes led to a higher incorporation percentage of the fluorinated amino acid 4-FT in proteins of both mutants. The mutant lacking FthB did not show any hindrance in growth but its expression in a heterologous streptomycete did increase its resistance to 4-FT. On the other hand, the strain lacking FthC showed higher accumulation of 4-FT intracellularly and greater sensitivity when added 4-FT in the growth medium [48].

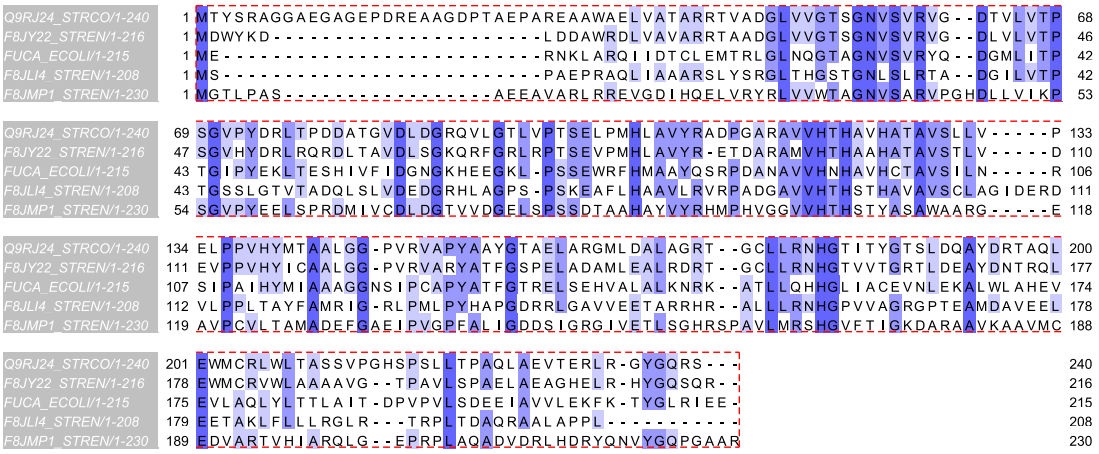


Figure 2.4: Alignment of 3 aldolases from from *S. cattleya* and fuculose aldolases from *S. coelicolor* and *E. coli*.

2.3.2 The fluorinase enzyme

The first step in the pathway of 4-FT synthesis, as illustrated above is realized by 5'-FDA synthase (or fluorinase), the only known enzyme to date that is capable of catalysing the incorporation of inorganic fluoride to biologically relevant organic molecules. This enzyme has enjoyed much attention in the following years both because of its uniqueness in catalysis of direct fluorination reactions combined with the biological importance of fluorinated chemistry. Immediately after its discovery the crystal structure has been solved [49] confirming the hexameric conformation as a dimer of trimers.

The mechanism of action that has been proposed is an S_N2 nucleophilic substitution with the fluoride anion acting as the nucleophile [50]. Initially, the water solvated fluorine atom at complex with 4 water molecules is passively diffused into the binding pocket and binds to the 2 hydrogens of serine 158 residue of the fluorinase displacing 2 water molecules. SAM then sits upon fluoride expelling the 2 remaining water molecules and initiating the nucleophilic attack to the 5-carbon of the ribose ring substituting sulphur with fluoride. The proposed mechanism of

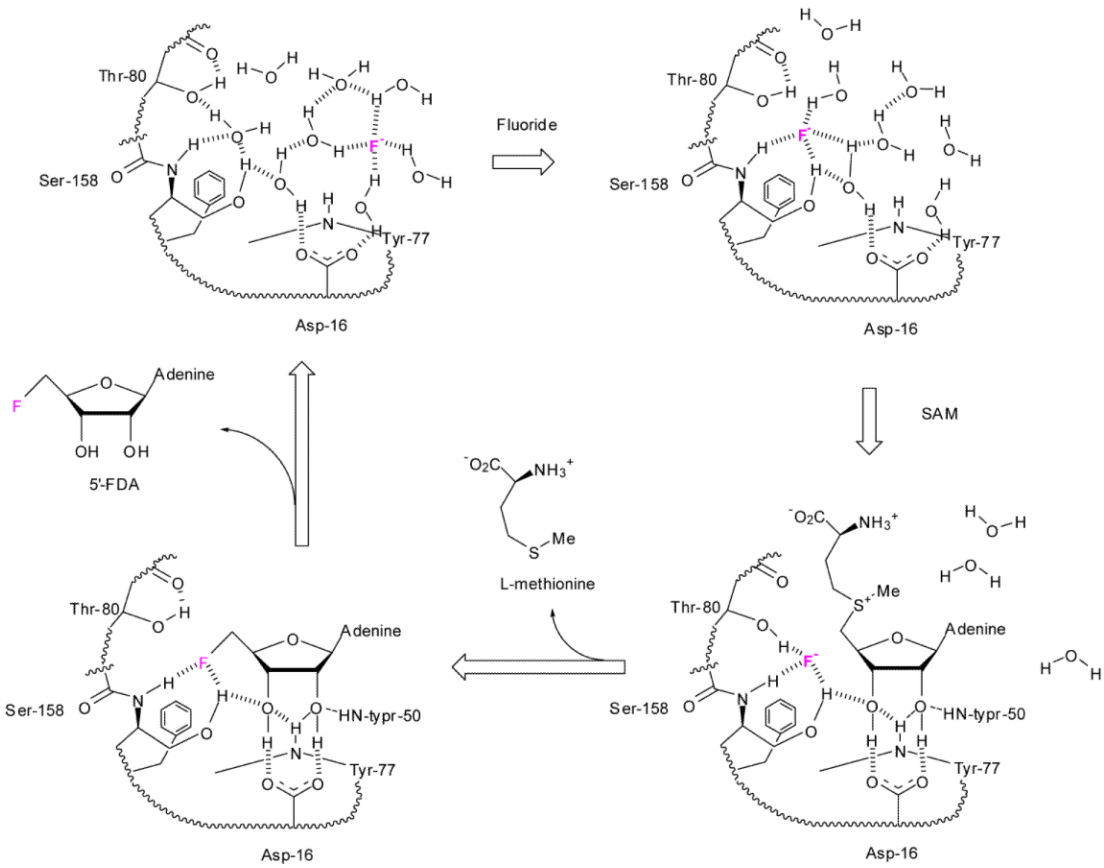


Figure 2.5: The mechanism of action of the fluorinase [3]

action was also strengthened by an experiment where the enzyme was allowed to equilibrate with one of the substrates first before addition of the second. When SAM was added first, the reaction showed a 6-fold less burst in reactivity and therefore active-site availability, indicating that the SAM-bound enzymes show lower affinity for fluoride presumably because the accepting pore is shadowed [3]. The substrate specificity of the enzyme was also an item of research in subsequent studies for the identification of potential inhibitors and alternative functionality. Two molecules similar in structure with SAM were initially tested, both of which are established inhibitors of SAM-dependent enzymes. First, a by-product itself of those enzymes, S-Adenosyl-L-Homocysteine which results from the cleavage of a methyl-group from the sulphur moiety and second, sinefungin which differs from SAM in that sulphur is replaced with a carbon connected to an amine group. It

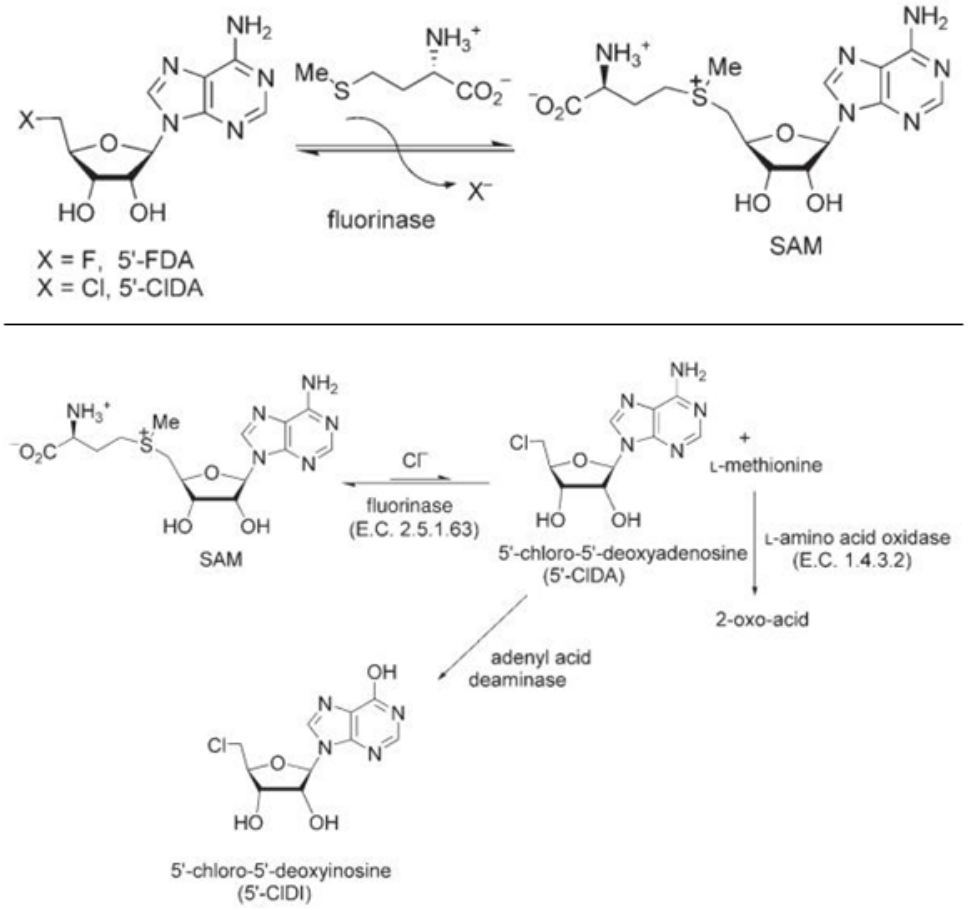


Figure 2.6: The broad halogen specificity of fluorinase (top) and the combined enzyme assay revealing the chlorination mechanism (bottom) [4].

has been shown in assays that contain equal concentrations 0.4 mM of both SAM and *S*-adenosyl-L-homocysteine (SAH) (10 mM KF, 7.8 pH, 37 ^{o}C), that the enzyme loses 95% of its activity, therefore SAH is a strong competitive inhibitor of the fluorinase with a $K_i = 29$ µM [41]. Furthermore, in another study is has been shown that SAH binds to the enzyme with 40-fold higher binding affinity than SAM [3]. The latter was an item of concern for the development of *in vivo* fluorination in *E. coli* presented in chapter 4, as SAH is also present in *E. coli*, albeit in low molarities (1.3 µM reported in [51]) as a result of the action of MTA nucleosidase which is part of the methionine salvage cycle. Sinefungin, on the

other hand showed only weak inhibition with 15% loss of activity in equimolar solutions of SAM and sinefungin.

The fluorinase has naturally been studied for substrate affinity with other halogens. Initially, assays with chloride failed to reveal any production of the chlorinated analogue, 5'-ClDA [41]. However, a later study has shown that the absence of 5'-ClDA was in fact a result of the equilibrium of the reaction heavily shifted towards the substrates. With an ingenious approach, including two more enzymes that irreversibly convert the products of the chlorination reactions, namely 5'-ClDA and L-methionine to 5'-chloro-5-deoxyinosine (5'-ClDI) and 2-oxo-acid respectively, the fluorinase was proven to possess the capacity to function as a chlorinase. Moreover, in the same study it has been shown that halogenation with the fluorinase is a reversible reaction and that this reversibility combined with the preference of fluoride over chloride can be used in a two-step transhalogenation reaction with the fluorinase for conversion of 5-ClDA into 5'-FDA [4]. The only directed evolution study involving the fluorinase that has been reported so far had this two-step reaction as an optimization objective, though native fluorination activity was hindered in the isolated variants [52].

2.3.3 Other fluorinase variants

Recently, 10 years after the discovery of the first fluorinase from *S. cattleya*, 3 more variants have been identified by genome mining [53]. They all showed high sequence homology (80%-87%) with the fluorinase of *S. cattleya* and among themselves, and are almost structurally identical, while there is a 21 consecutive amino acid region which comprises the active site of the enzyme and is conserved throughout a family of proteins that catalyse the conversion of SAM to adenosine analogs. The bacterial hosts accommodating the fluorinase variant genes are *Nocardia brasiliensis, Actinoplanes sp.* and *Streptomyces* sp. MA37. Recombinant

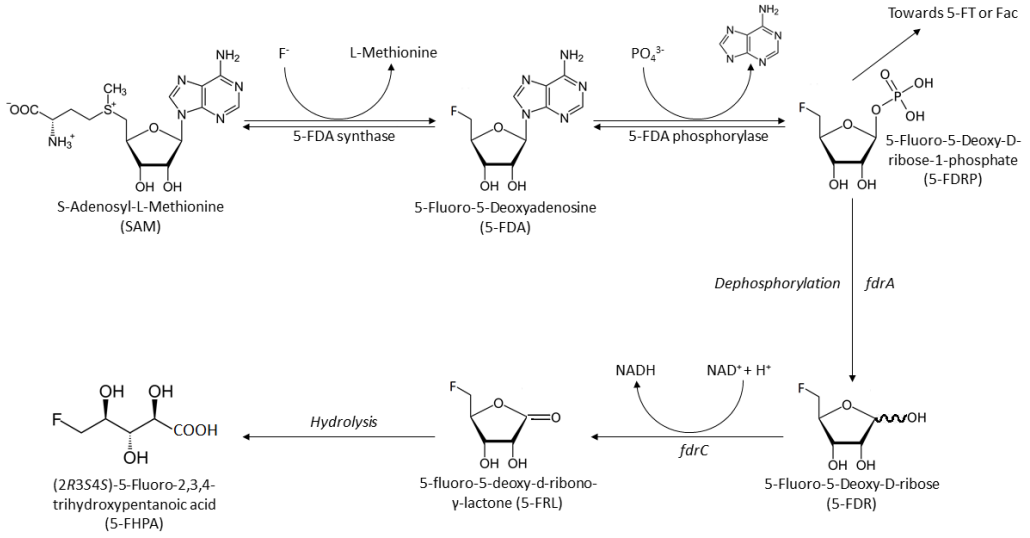


Figure 2.7: The alternate route of the fluorometabolite pathway in *Streptomyces* sp. MA37 involving enzymes *fdrA* and *fdrC*.

versions of these enzymes were synthesized and overexpressed in *E. coli*. Assays were performed and all these three variants showed higher catalytic efficiency from the one in *S. cattleya* with most efficient the one from *Streptomyces* sp. MA37. Apart from the fluorinase homologues, the other pathway enzymes have also been tested for homology. In all 3 bacteria accommodating variants of the fluorinase, genes coding for putative pathway enzymes have also been found with more than 50% sequence identities compared to the *S. cattleya* homologues. The identification of these genes also raised the possibility of other potential fluorinated metabolites being elaborated by those bacteria. In *N. brasiliensis* incubated with fluoride, no fluorometabolites were detected and therefore the pathway was deemed to be inactive. No assays were performed with *Actinoplanes sp.* as it is not available in the public domain. The case of *Streptomyces* sp. MA37 is an interesting one as it not only elaborates fluorometabolites, but also accommodates an alternative route in the pathway leading to a novel fluorometabolite (2*R*3*S*4*S*)-5-fluoro-2,3,4-trihydroxypentanoic acid (5-FHPA) as shown in figure 2.7. The two

enzymes employed in the conversion of 5-FDRP to 5-FHPA are products of *fdrA* and *fdrC*. Both of those genes were discovered by sequence homology with the *Salinispora tropica* pathway (more details in the next section) fuelled by similarity in the first two steps with the fluorothreonine pathway. Specifically *salN*, the product of which is able to dephosphorylate 5-ClDRP to 5-ClDR in salinosporamide synthesis, showed a 56% homology with *fdrA*, therefore it was assumed that FdrA is the enzyme mediating dephosphorylation of 5-FDRP to 5-FDR. The second step, found to be catalysed by the product of *fdrC* (69% identity with *salM*), is the oxidation of 5-FDR yielding 5-fluoro-5-deoxy-D-ribono-γ-lactone (5-FRL) which subsequently yields 5-FHPA by hydrolysis [54]. In another study, it was discovered that the marine bacterium *Streptomyces xinghaiensis* was able to elaborate FAc in a sea salt-dependent manner [55]. This finding led to the discovery of another variant of the fluorinase gene in the genome of this bacterium, following a the same trend of very high sequence identity and structural similarity with the first fluorinase (84%). Kinetic studies showed that this variant shows marginally higher catalytic efficiency compared to the most efficient fluorinase previously characterized (FlA1 from *Streptomyces* MA37 - $k_{\text{cat}} = 0.277$ min^{-1} instead of 0.260 min^{-1}), but surprisingly higher specificity for SAM, more than 10-fold compared to FlA1 [1]. A total of 5 variants of the fluorinase have therefore been characterized so far and their Michaelis - Menten rates are collectively shown in table 2.1. The more recently discovered fluorinase (FlA4) exhibits a $K_m = 8.2$ mM for fluoride as determined with the same assay, but with a steady 500 μM concentration of SAM and increasing fluoride 0 - 30 mM.

Except for the above characterized fluorinases, a BLAST search on the NCBI database as of 20/11/2018, produces 4 more ORFs which are predicted to encode for fluorinases with a sequence identity of more than 50%. Specifically, two sequences, one from the *Actinopolyspora mzabensis* genome and one from *Amycolatopsis* sp. CA-128772 exhibit identities of 78% and 81% respectively. One

Fluorinase source	V_{max} (µM min^{-1})	SAM K_m (µM)	$k_{cat} = V_{max}/C$ (min^{-1})	Specificity constant = k_{cat} $/K_m$(mM^{-1} min^{-1})
Streptomyces xinghaiensis (FlA4)	3.88 ± 0.11	7.04 ± 0.94	0.277 ± 0.007	39.5 ± 1.51
Streptomyces sp. MA37 (FlA1)	3.63 ± 0.13	86.0 ± 11.3	0.260 ± 0.004	3.02 ± 0.19
Actinoplanes sp. N902-109 (FlA3)	2.78 ± 0.15	43.1 ± 7.99	0.197 ± 0.003	4.58 ± 0.50
N. brasiliensis (FlA2)	1.77 ± 0.07	30.4 ± 4.24	0.128 ± 0.004	4.22 ± 0.42
Streptomyces cattleya (FlA)	1.17 ± 0.06	29.4 ± 5.80	0.084 ± 0.005	2.84 ± 0.14

Table 2.1: The Michaelis - Menten kinetic constants for the 5 variants of the fluorinase as measured in [1]. The kinetic rates were determined with an assay of increasing SAM from 0 to 800 µM in a saturating concentration of fluoride (200 mM). The concentration of protein was 14 µM in 20 mM of sodium phosphate buffer (7.8 pH) and a temperature of 37^oC.

comes from *Peptococcaceae* bacterium CEB3 which has a 66% identity and automatically annotated as a chlorinase, and finally one from *Thermodesulforhabdus norvegica* with a 55% identity. The alignments are shown in figure 2.8. It would be surprising if at least the ones with high homology are proven not to catalyse the well-known fluorination reaction. The latter is also strengthened by the fact that ORFs with homologues of *flB* (MTAP) exist with sequence identities 59% for *A. mzabensis* and 53% for *Amycolatopsis* sp. CA-128772. The same is true for the 5-methylthio-ribose-1-phosphate isomerase but with somewhat lower identities (51% and 43%) respectively. For the aldolase enzyme, one out of two putative

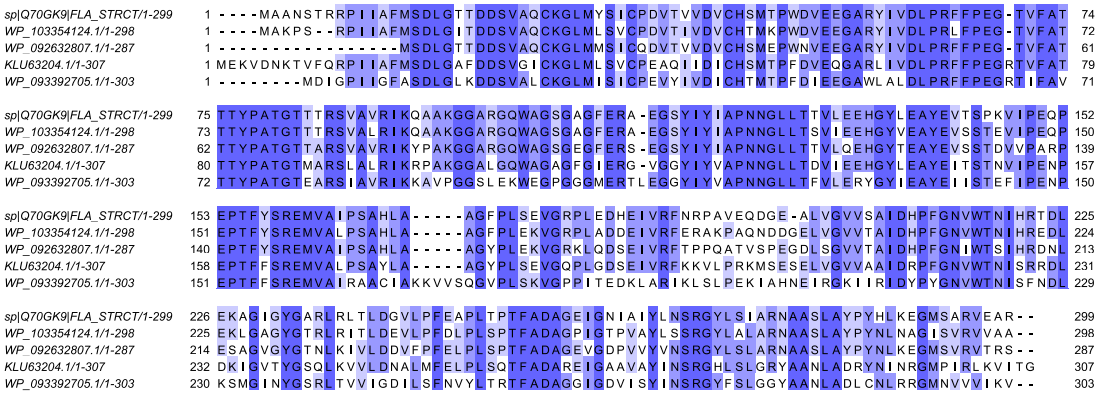


Figure 2.8: Four ORFs encoding putative fluorinases from *Amycolatopsis* sp. CA-128772, *A. mzabensis*, *Peptococcaceae* bacterium CEB3 and *T. norvegica*.

aldolases of *S. cattleya* has a high identity homologue (59%) in *Amycolatopsis* sp. CA-128772. Finally, homologues of the 4-FT transaldolase that performs the last step of the pathway exist in both *A. mzabensis* (62%) and *Amycolatopsis* sp. CA-128772 (64%). Evidently, even though the current convocation of fluorinases is quite poor, with more and more sequences being deposited from newly discovered organisms, more potential fluorinases will appear.

2.3.4 Engineered *in vivo* fluorination

The only instance of engineered direct fluorination in another living bacterium is fluorosalinosporamide production which was made possible by simply substituting the *salL* gene with the fluorinase (*flA*) gene [5]. These two genes exhibit the highest "non-fluorinase" sequence identity of 40% and they catalyse virtually the same reaction with identity of the halogen being the only difference. Presumably, the deletion of the *salL* gene and substitution with *flA* came as a natural choice for investigating the compatibility of the fluorinated analogues in the next steps of the pathway towards salinosporamide A and they were found indeed to be compatible yielding the fluorinated version of salinosporamide A. The steps are

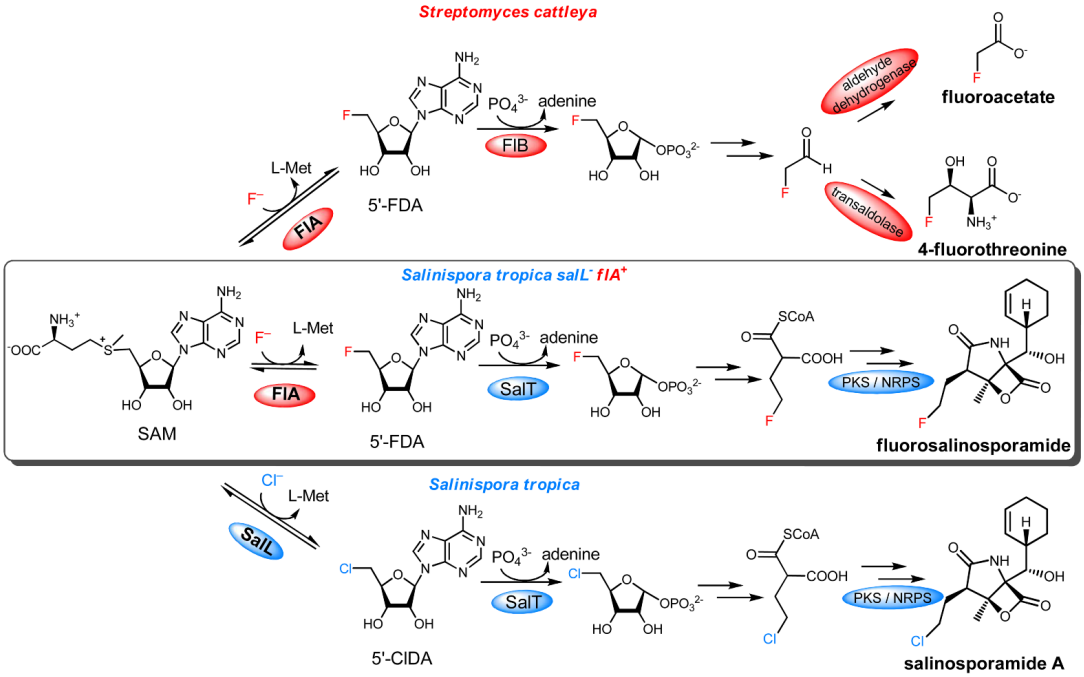


Figure 2.9: Substitution of fluorinase in the pathway for salinosporamide A biosynthesis in *S. tropica* [5].

shown in figure 2.9. Although the latter was elaborated with a 10-fold lower yield than the wild type yield of salinosporamide A, it was the major product of the pathway among 5 fluorometabolites that were detected. The identity of those was not elucidated however it is thought that they are close analogues of fluorosalinosporamide as they exhibit similar coupling patterns.

A number of studies reveal the implementation of *in vivo* fluorination pathways in other broad usage organisms such as *E. coli*. An interesting case is the production of 2-fluoro-3*R*-hydroxybutyrate (FHB) through the *in vivo* activity of an engineered pathway beginning with fluoromalonate, a non-toxic analog of FAc, by expressing a malonate transporter protein for intracellular delivery of the fluoromalonate, proceeding with a conversion to fluoromalonyl-CoA by a malonate:CoA ligase from *Rhodopseudomonas palustris*, continuing with production of acetofluoroacetyl-CoA by action of an acetoacetyl-CoA synthase and finally

production of 2-fluoro-3*R*-hydroxybutyryl-CoA by an acetoacetyl-CoA reductase. The latter was then either reduced to FHB by hydrolysis from native processes or polymerization to fluorinated poly(hydroxyalkanoate) bioplastics. The amounts of end-products were lying in millimolar levels after 48h of incubation [56].

2.4 DNA editing techniques

The ability to alter the DNA of any chosen organism from the simplest virus up to the higher mammalian cells is not only essential for the study of the organism itself and its genetic elements but also a field-defining feature for SB. The precision of DNA editing and reduction of off-site or polar effects from genome editing that are not tied to the presence of the gene product itself is of outermost importance for the legitimacy of research findings. Fortunately, the advancement of available tools for gene deletion, insertion or modification along with the highly developed sequencing technology, have made the accurate editing of the genomes of many organisms routinely achievable and less time-consuming. There is a wide variety of DNA editing techniques, each of them entailing a large number of variants and most of them are tailored specifically to the organism where editing takes place. In the following sections, methods for altering the genetic environment of the *E. coli* bacterium will be described.

2.4.1 Gene editing by native recombination machinery

E. coli accommodates a system which comprises by set of proteins induced when DNA damage is detected in the bacterium by appropriate control elements. The proteins are responsible mainly for DNA repair and their collective function is called SOS response. Part of this system is the recombination which promotes repair of damaged gene by replacing with a homologue found "nearby". There

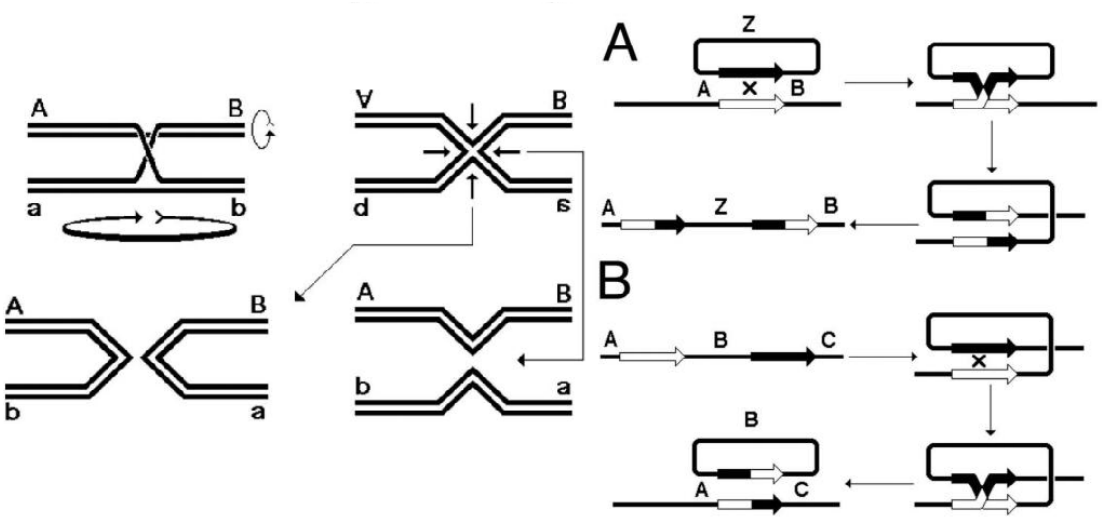


Figure 2.10: Left: Formation of a single Holliday junction and resolution thereto. Right: Insertion (A) and excision of a plasmid (B). [6].

is a large number of proteins that are involves in the recombination mechanism but the key proteins are called RecA and RecBCD. Defects in these genes have serious repercussions in cell viability with or without induced DNA damage [57]. RecA is particularly important as it creates a complex with ssDNA created from a probability driven Double Strand Break (DSB) and sweeps through the chromosome until it finds a large region of homology, usually more than 500bps. At this point the ssDNAs created from the DSBs are paired with the other sequence and a "Holliday Junction" is created as shown in figure 2.10 - Left. This native recombination ability can be "tricked" to replace specific DNA sequences in the chromosome with sequences of choice. In a revolutionary method released 30 years ago, it was shown that by designing a plasmid which harbours a temperature-sensitive origin of replication, and accommodates a mutant allele of a gene to be edited including regions of homology upstream and downstream, it is possible for this plasmid to be integrated in the chromosome on the 5' or 3' end of the gene by utilizing the innate homologous recombination system. Specifically, the RecA protein is enabled and creates a holliday junction which if resolved properly results in the insertion of plasmid in the chromosome as shown

in figure 2.10 - Right. Clones which have managed to integrate the plasmid are called cointegrates and they are usually selected in LB agar plates with antibiotic whose resistance is encoded in the plasmid and a temperature which is not permissive for the normal replication of the plasmid, therefore the surviving clones are cointegrates. Integration events are rather rare with a frequency of 10^{-3} - 10^{-4} because they depend on a DSB happening well within the region of homology in either strand, a high number of transformants have to be plated for adequate clones to appear [58].

2.4.2 Gene editing by foreign recombination

Recombineering refers to genetic engineering (i.e. DNA editing) using homologous recombination techniques that are mediated by recombinant proteins from phages. Although, genetic engineering using phages is rather old, techniques have been improved first by removing elements which cause virulence, thereby permitting the stable integration of a prophage in the chromosome, and further developed with isolation of the minimum essential genes that enable recombination functionality. There are many protocols and variants of this method, however it is based on three proteins from lambda phage, Exo, Beta and Gam. The essential materials consist of a tightly controlled expression system for this set of proteins and the intracellular presence of a double or single stranded DNA (dsDNA or ssDNA respectively) with tails of around 50 bps, homologous to the target sequence. In fact, the word "recombineering" itself has been invented after the first implementation of the method with dsDNA, by a study which improved upon the method using ssDNA instead [59]. Exo and Gam operate in concert and perform the actual recognition and recombination while Gam is responsible for the suppression of RecBCD, an exonuclease which cleaves linear DNA that is present in an *E. coli* cell [60].

Recombineering is still the most frequently used method in gene editing involving research in *E. coli* and an attempt at description of the steps involved will be attempted in a generic format below. The keen reader can refer to [61] for a more detailed description.

Step 1) Choice of the DNA sequence to be edited (deleted, substituted or inserted).

Step 2) Obtain the template for editing. This can either be created by simple PCR, fusion or mutagenic PCR from the source harbouring the foreign DNA, or obtained directly as synthesized DNA. In these days the latter is a very attractive choice as a 1000bp DNA sequence can be synthesized for as little as 100 pounds. The template can be either a selection marker (e.g. antibiotic resistance cassette) or in case of insertion of a foreign gene or modification of an existing gene, the modified (or to be inserted) sequence or both.

Step 3) Creation of appropriate primers which apart from the complementary base pairs with the template carry short (30-50bp) flaking homology arms with the sequence upstream and downstream of the DNA to be edited. Amplification of the template by PCR using the aforementioned primers. The end product is the initial template flanked with the homology arms.

Step 4) Transformation of the *E. coli* cells with a plasmid containing 3 genes for the lambda red functions under the control of a temperature sensitive cI-repressor. Growth of the transformed strains in 30^{o}C - 32^{o}C up to mid-log phase and then placement in a 42 ^{o}C warm water bath for 15 minutes with a 200 RPM rotary shaking. This is a pulse-like induction of the lambda phage recombination machinery. Afterwards, the culture has to be put in a ice cold water bath in order to restore repression and freeze the state of the cell with the expressed proteins.

Step 5) Transformation of the cells from step 4 with the linear DNA with the homology arms obtained from steps 1-3. Incubation of the culture in (30^{o}C -

32^{o}C) and then streak in LB agar plate containing the antibiotic for overnight incubation. Selection of surviving colonies with the inserted fragment from the plate.

There are several advantages of this technique. Apart from its relatively quick implementation time, it can be used in *E. coli* strains with impaired native recombination machinery such as recA1, recA13 genotypes typically seen in cloning strains.

2.5 Molecular biology

The techniques that have been consistently utilized in this work, for bacterial maintenance, modification and handling are presented below.

2.5.1 Media preparation

E. coli bacteria are grown depending on the type of study in either rich or minimal medium. LB (Lysogeny Broth) is routinely utilized as a rich medium for *E. coli* and contains yeast extract (5 g/L), tryptone (10 g/L) and NaCl (10 g/L). All cultures in this work were grown on LB unless stated otherwise. Minimal medium for *E. coli* (M9) contains the absolute essentials for *E. coli* growth: a carbon source (glucose or glycerol 0.4%), $MgSO_4$, $CaCl_2$ and M9 salts ($Na_2HPO_4.7H_2O$, KH_2PO_4, NaCl and NH_4Cl). Solid media are used in petri dishes for plating bacteria and isolating single colonies. Both LB and M9 can be solidified by adding agar in the above mixtures.

2.5.2 Gel Electrophoresis

Gel electrophoresis involves separation and visualization of DNA molecules according to their size. This technique is based on the electro-negativity of DNA molecules and when placed inside an electric field, they are attracted towards the positive pole. The first step involves creation of the agarose gel, a solidified substance with wells placed at the top. In this study, all gels were created with TAE buffer (Tris base, Acetic acid and EDTA) adding 1% agarose and SYBR^{TM}Safe DNA Gel Stain (Thermo Fisher Scientific^{TM}). The apparatus is very simple and includes a container with two electrodes, one on each end. The gel is placed into the container which is filled with TAE buffer until the gel is completely submerged into it. The samples to be analysed are placed in the wells and power cables are attached to the electrodes. Then current ($\approx$ 100 mA) is left to pass through the TAE buffer and the gel for at least 30 minutes. The smaller sized DNA molecules run faster through the solidified gel, thus DNA bands can be resolved according to their size. Usually, one or two wells at the edges are reserved for a DNA ladder which contains DNA of predefined sizes serving as a rule for the rest of the samples. Finally, the bands are visualized by exposure to UV light which illuminates the DNA stain attached to the molecules.

2.5.3 Plasmid extraction

Plasmid extraction is the process of acquiring ample quantities of plasmid DNA from lysed *E. coli* bacterial cells. In this work, all plasmid extractions were performed with the use of a *GeneJET Plasmid Miniprep Kit* (Thermo Fisher Scientific^{TM}). This kit is an adaptation of the widely utilized alkaline extraction procedure in which chromosomal (but not plasmid) DNA is denatured in a solution with high alkalinity and then renatures and becomes insoluble upon neutralization [62]. The process is as follows:

- Growth of a liquid culture to saturation with the appropriate antibiotic to ensure plasmid maintenance.
- Centrifuge the culture tubes in 6,000g for 5 minutes. Discard of supernatant (S/N). After the procedure, bacterial cells are in the form of a pellet located at the bottom.
- Resuspend the pellet in appropriate buffer. This buffer also contains RNase in order to degrade RNA. The solution is now over-saturated in bacterial cells and completely opaque.
- Add lysis buffer (high alkalinity) and invert the tube gently 5-6 times. The solution becomes transparent and appears viscous.
- Add neutralization buffer (low alkalinity) and invert the tube 5-6 times. The renatured chromosomal DNA appears as white clot in the solution.
- Centrifuge for 5 minutes in >12,000g to pellet cell debris and chromosomal DNA. Plasmid DNA resides the S/N.
- Transfer of the S/N on the specially designed GeneJET column for the isolation of plasmid DNA.
- Spin down the solution (1 minute in >12,000g). DNA is trapped in the column.
- Add wash solution and centrifuge for 1 minute (>12,000g) to clean the column of unwanted remaining debris (repeat 2 times).
- Finally, DNA is eluted by addition of an elution buffer (mostly water), incubation for 2 minutes and centrifuging in >12,000g for another 2 minutes.

It is important to note that plasmid maintenance is done in specific *E. coli* strains which are deficient in DNAses (such as DH10B) and recombinases, otherwise

residual DNAse activity may degrade the DNA after elution in the tube and incubation in 37 ^oC. If a plasmid needs to be extracted from a strain containing DNAses, further purification with a PCR cleaning kit (see 2.5.4) can ameliorate the problem.

2.5.4 DNA assembly

Construction of plasmids presented in this work has been achieved with the use of FastDigest^{TM}restriction enzymes (Thermo Fisher Scientific^{TM}) and the T4 ligase (NEB). After each reaction, DNA was subjected either to gel electrophoresis (2.5.2) followed by excision of the desired band and purification with the use of a Wizard® SV Gel and PCR Clean-Up System (Promega), or directly to purification with the same kit.

2.5.5 Chemical transformation

Transformation is the process of DNA uptake from bacterial cells and while there are species that already posses the ability to uptake extracellular DNA, termed natural competence, *E. coli* does not belong in this category. However, competence can be induced by chemical treatment of non-competent bacteria with highly concentrated $CaCl_2$. Calcium anions are thought to create pores in the membrane of *E. coli* and DNA uptake is then achieved by incubating the chosen DNA together with the bacteria [63]. The anions bind both in DNA molecules and lipids in the membrane owing to them both being electronegative, thus bringing them together and enabling the attachment of DNA in the cell surface, which otherwise would be difficult due to mutual repulsion. This procedure works better in a low temperature because the fluidity of the cell wall is reduced. The mixture

is then subjected to a thermal shock which creates thermal currents to the pores in the membrane enabling the uptake of DNA.

2.5.6 Polymerase Chain Reaction

PCR is one of the most utilized tools in experimental molecular biology and constitutes a considerable chunk of the underlying experiments in the present work. This method has been made possible by the discovery, 30 years ago, of a thermostable DNA polymerase in an extremophile bacterium *Thermus aquaticus (Taq)* [64]. In short, starting from a DNA sequence which can be either linear or circular (template), PCR mediates the *in vitro* copying of this initial DNA in many rounds to finally yield a greatly amplified amount ($2^{30} - 2^{40}$) of this particular sequence.

The materials needed for the PCR process are:

1) An initial DNA template sequence.

2) A thermostable DNA polymerase.

3) Two primers. Short DNA oligonucleotide sequences (18-22 base pairs) to bind in the primary DNA strand (forward primer) and the complementary strand (reverse primer).

4) Single deoxynucleotide triphosphates of four types (dATP, dTTP, dGTP, dCTP) which correspond to the four DNA bases A, T, G and C and are the substrates accepted by the DNA polymerase to perform elongation of the DNA.

5) A suitable reaction buffer

The reaction requires an thermoblock able to robustly control the temperature and includes the following steps:

1) Denaturation of the template DNA which is done in 95-98 ^oC. In these temperatures the hydrogen bonds between complementary bases break resulting in single stranded DNA.

2) Annealing of the primers to the single stranded sequences from step 1 (58 - 62 ^oC).

3) The DNA polymerase binds to the double stranded ends with the primers and starts the extension process of the complementary DNA strands, finally yielding two copies of the initial DNA sequence.

4) Repeat from step 1.

The above cycle is repeated 20-40 times and the number is based on several factors, such as the desired final quantity of DNA template and the efficiency of the reaction which in turn is influenced by the quality of the template and its accessibility by the polymerase, the primer sequences and the specific DNA polymerase used. Most issues arising in PCR are related to the choice of primers. These include secondary structure formation, non-specific binding and primer dimerization, where the forward primer binds to the reverse primer. There are specific workarounds to alleviate the detrimental effects such as tweaking the temperature or duration of particular steps, addition of specific chemicals that prevent secondary structure formation or enhance polymerase efficiency, or choice of another polymerase. However, there are tradeoffs most of the time. Selecting the appropriate conditions for a successful PCR is sometimes challenging and this procedure is termed PCR optimization.

Primer design is the process of specifying two exact nucleotide sequences which are essential part of step 2. Most of the time, biologists use a simple formula to specify the primer sequence which is *T = 4 x (Number of G or C bases) + 2*

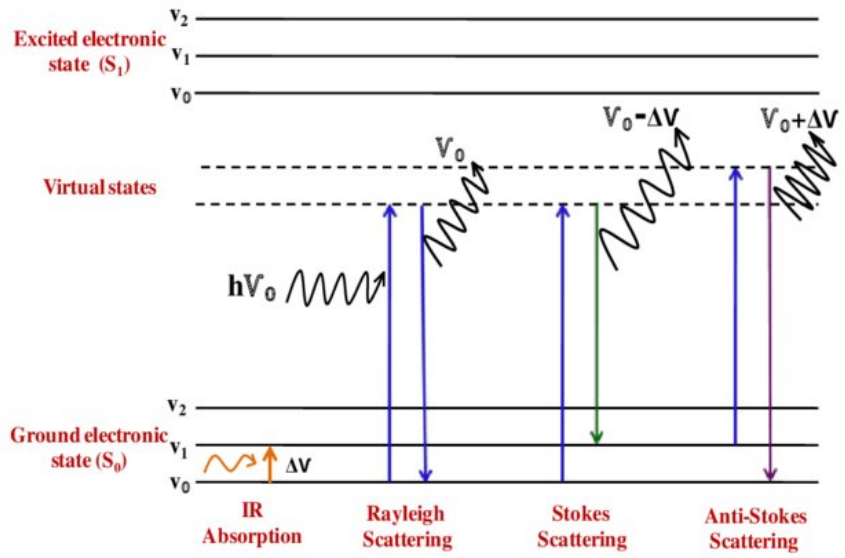


Figure 2.11: Types of scattering as illustrated in [7].

x (Number of A or T bases). This simple formula was routinely used in primer design as part of this work.

2.6 Raman Spectroscopy

RS has proven to be a powerful diagnostic tool that has been used during the past decade in biomedical applications and disease diagnosis with novel studies being produced at an unprecedented rate [65]. An illustration of Raman scattering is shown in figure 2.11. According to Plank, photons are undivided "packets" of energy which is determined by their wavelength (λ), or frequency (ν), as $E = hc/\lambda$ or $E = h\nu$. In principle, if the photonic energy is within an appropriate window, photons that "hit" a molecule (incidence photons) can be absorbed, exciting the energy state of atomic (UV range) or molecular bonds (IR range). The latter is the basis of IR spectroscopy. If a photon is not absorbed, the other possibility is that it is scattered. There are 3 types of scattering: Rayleigh, stokes and anti-stokes. Most of the scattered photons emitted by the irradiated substance are of Rayleigh

type, and have the same wavelength as incidence photons. A tiny fraction (10^{-7}), is of the other two types. RS measures the Stokes photons generated, in which case the photon excites the molecule to a "virtual" energy state but some of this energy is absorbed to a molecular bond as real energy, which is exited to a higher energy state. The scattered photon loses an equal amount of energy, therefore appearing shifted to higher wavelengths. Based on the difference of these shifted photons the types of molecular bonds of a sample can be elucidated. This is the basis of RS.

The Raman spectra of bacteria have been extensively used in research to distinguish between different bacterial species which normally exhibit varied spectra with respect to one or more metabolites [66]. There are Raman setups that highly enhance the limit of detection by up to five orders of magnitude [67]. In particular, it has been shown that coherent Raman methods such as Coherent Anti-Stokes Raman Spectroscopy (CARS) and Stimulated Raman Spectroscopy (SRS), exhibit an increase in signal of at least 2 orders of magnitude in comparison to spontaneous Raman scattering when measuring bacterial endospores [68].

Chapter 3

Bacterial growth curve modelling and promoter quantitation

SB as a field is based on the mechanistic view of a biological system, and as such, accurate quantification of the interactions between the basic elements that comprise this system is vital. There is a large variety of regulatory elements that constitute several discrete categories according to the type of regulation. Some of these categories are discussed briefly in section 2.1.2.

Promoters in particular have been receiving most of the attention, mainly because they are the most commonly occurring regulatory element, however this does not diminish the importance of other regulatory elements, especially when viewed under the light of novel regulatory network construction. The rate at which novel promoters are discovered or synthesized is unprecedented, whereas the characterization procedure still suffers from a few bottlenecks, as a novel element has to undergo experimental validation across different hosts and conditions in order to reliably demonstrate a robustly predicted behaviour. The latter indicates the need for a high-throughput method in the characterization of novel elements.

The approaches to experimental characterization can be categorized to direct and indirect. An example direct approach is the real time quantitative PCR which can measure the amount of transcript for a particular gene modulated by a regulatory element. This approach, whilst being sufficiently accurate, suffers in terms of speed because the procedure requires "destruction" of the sample. A popular indirect approach for characterization involves Fluorescent Protein (FP). These proteins, when folded (matured), they reflect light in particular wavelengths. The number of photons reflected can be used as a measure of the amount of proteins that currently reside in the sample. When the gene corresponding to an FP is tied to a particular promoter (or other regulatory element/network) it yields a measure of its translational activity. While this approach suffers in accuracy compared to its counterpart, it makes up for speed, because it enables the continuous monitoring of a growing bacterial culture.

The majority of studies involving experimental validation of regulatory elements with FPs consistently represent promoter activity as the output of the fluorescent signal normalized to the biomass, which is given by the OD of the culture. The experimental values are usually obtained from a coupled time-series experiment. In the following sections, a case will be made for the need of a unified metric which takes into account several aspects of FPs and the growth of bacteria in a liquid culture with limited resources. Following the application of simple mathematical reasoning, an implemented method that describes promoter activity at each time point will be illustrated. Protein expression and maturation, are modelled as first-order differential equations, taking into account the degradation and maturation rates of the FPs which need to be known in advance. The promoter activity is then expressed based on the measured values of fluorescence and OD with a formula derived by mathematical manipulations of the defined quantities and the differential equations that comprise the model. Continuous expressions for fluorescence and OD are obtained from *GP* regression. Validation of the tool with experimental data from several constructs reveals the potential to effectively

quantify promoter activities in a time-series manner regardless of the host and the FP used.

3.1 Mathematical modelling of bacterial growth curves

In this section, the traditional and latest approaches for the modelling of bacterial growth are presented including various informative measures that can be extracted from the growth curves. A detailed analytical description will be given for the novel approach of using *GP*s for fitting growth curves and an improvement of this presented in the paper will be highlighted.

Bacterial growth in limited resources has been an object of interest since Monod's work on growth of bacterial cultures over 60 years ago [69]. Monod adapted a simple concept, correlating the consumption rate of an essential nutrient with the rate of growth. According to Monod, the rate of bacterial growth (μ) obeys the following equation:

$$\mu = \mu_0 \frac{s}{k_s + s} \tag{3.1}$$

where s denotes the current substrate concentration, k_s is a constant assigned to the substrate concentration where the rate of growth is halved and μ_0 is the maximum rate occurring in saturating substrate ($s >> k_s$).

3.1.1 The Gompertz model

Underlying bacterial growth rate, the direct monitoring of the population of bacteria permits the experimental measurement of bacterial growth. The OD of a liquid culture is a widely used metric for accomplishing this. The resulting time-series of measurements provide the ability to directly model the growth rate and assign values to it as a function of time. It has been made evident that most bacteria growing in a liquid of a predefined volume comprising of the appropriate nutrients, follow a sigmoidal trend, indicating that the Monod's law takes effect. There are several approaches to the modelling and fitting of growth curves. A summary of these can be found in [70]. The mostly utilized model for bacterial growth is the Gompertz model:

$$G(t) = K \cdot exp[-exp[\frac{\mu \cdot e}{K}(\lambda - t) + 1]] \tag{3.2}$$

The parameters (K, μ, λ) correspond to the maximum growth (K, which is also denoted as the carrying capacity of the medium), the growth rate during the exponential phase (μ) and the lag time (λ). These parameters describe the three stages observed during the bacterial growth. Other simpler models, such as the logistic function (2 parameters), fail to capture some of the essential aspects of growth such as the lag time, whilst other more complex, such as the Richards or Stannard models (4 parameters) do not contribute significantly in the accuracy of model fitting, while still remaining rigid to divergent behaviours during growth. This is the reason why one of two methods presented in this chapter utilizes the Gompertz model. The process of fitting of a time-series of OD measurements to the Gompertz model yields an $G(t)$ as a function of time where the parameters (K, μ, λ) have been assigned specific values. The fitting is based on the minimization of the mean root squared error, and is performed from a multi-parametric constrained minimization algorithm provided in MATLAB^{TM}.

The resulting growth function $G(t)$ can be differentiated to yield the rate of growth as a function of time. The derivative is:

$$\frac{dG(t)}{dt} = \mu \cdot exp[-exp[\frac{\mu \cdot e}{K}(\lambda - t) + 1] + \frac{\mu \cdot e}{K}(\lambda - t) + 2] \tag{3.3}$$

An important aspect of bacterial growth routinely mentioned in relevant research is the exponential phase of growth, defined as the window of growth where the bacteria are dividing rapidly in an exponential manner. While it is widely accepted that the exponential phase involves an interval centred around the maximum growth, choices for the limits of this interval vary. The algorithmic approach utilized in the tool presented here, uses as starting value for the exponential phase the end of the lag phase ($t = \lambda$), and as the ending value the time t_m where $[\frac{dG(t)}{dt}]_{t_m} = \frac{\mu}{e}$.

3.1.2 Gaussian Process regression for OD

The multi-parametric models utilized for fitting bacterial growth curves mentioned above such as the simple logistic function or the more complicated Richards model correctly take into account the underlying biological processes governing bacterial growth in a limiting nutrient environment; however, due to their rigidity they fail to capture special cases, where particular events during growth alter proliferation rate. Moreover, when noisy data are produced during OD measurements, the above process of fitting, based on the root mean squared error are unable to take it into account, and therefore, produce erroneous growth curves. Both of the above occurrences are commonplace in experimental data involving biology.

Recently, it has been shown that a non-parametric *GP*-based method can be

effectively utilised for fitting bacterial growth curves and other biologically relevant time-series data. Swain et al. [71] showed that *GP*s perform naturally better when dealing with noisy experimental data, especially in the calculation of the first and second time derivatives. In this section, the approach of fitting based on *GP*s will be described, namely *GP* regression. Essentially, a *GP* is a distribution of functions rather than a strictly defined function. A *GP* is characterised by a mean function and a covariance function in an analogous manner that a (multivariate) Gaussian distribution is fully defined by a mean vector and a covariance matrix. The main idea behind *GP* regression is that a 'prior' *GP* is imposed, which is characterized by a mean $m(t)$ and covariance $k(t, t')$, then a regression process is utilized to compute a 'posterior' *GP*. A covariance function $k(t, t')$ gives the prior covariance value between two time points t and t'. Given some normality assumptions (Gaussian likelihood and Gaussian noise) the posterior is also a *GP*, whose mean and covariance can be calculated analytically. The posterior mean and covariance functions are given in 3.4 and 3.5 correspondingly.

$$E[f^*] = K(T^*, T)[K(T, T) + \sigma^2 I]^{-1}\boldsymbol{y} \tag{3.4}$$

$$\boldsymbol{C}[f^*] = K^* - K(T^*, T)[K + \sigma^2 I]^{-1}K(T^*, T)^T \tag{3.5}$$

Let t_i be the ith time point of the experimental data and y_i be the value of the OD or fluorescence measurement at that time point. Let $T = [t_1, ..., t_N]$ be the training data time points and T^* the test data time points. A common choice for covariance function is the radial basis function defined as:

$$k(t_i, t_j) = \alpha^2 exp(-\frac{1}{2}\frac{(t_i - t_j)^2}{l^2}) \tag{3.6}$$

Then, the covariance matrix $K(A, B)$, where A and B are vectors of dimensions N and M respectively, is defined as:

$$A = \begin{bmatrix} k(a_1, b_1) & \dots & k(a_1, b_M) \\ \vdots & \ddots & \vdots \\ k(a_N, b_1) & \dots & k(a_N, b_M) \end{bmatrix} \tag{3.7}$$

The quantities α and l in the definition of the covariance function in 3.6 are termed amplitude and length respectively and along with the noise term σ, they represent the hyperparameters of the covariance function. Their definition can be performed based either on specific knowledge on the dataset or by performing an optimization step before the regression which only takes into account the noise term σ.

A key characteristic of the *GP* implementation is the ability to include two more types of information present in experimental data. First of all, any *ab initio* knowledge of the behaviour of the data can be included in the prior mean function, and hence calculation of the predicted mean function. Given a covariance function $k(t, t*)$, a *GP* can be described as:

$$f(t) = GP(m(t), k(t, t^*)) \tag{3.8}$$

It is a common practise to consider $m(t) = 0$ as the prior mean function, in which case the posterior mean is given by 3.4. This does not impose any limitations to the final predictive mean function, but equally does not contain any *ab initio* knowledge for inclusion in the model. In the case where there is specific prior knowledge, it is intuitive to include this within the prior mean function of the *GP*. This prior knowledge optimises the *GP* regression process which is now being applied to the residuals of the experimental data subtracted from the prior

curve. In the case of bacterial growth curves, the prior mean can be represented by the best fit of the Gompertz model 3.2 or any other specific model used to regress OD experimental data. Equation 3.4 is then modified, finally yielding the following:

$$E[f^*] = m(T^*) + K(T^*, T)[K(T, T) + \sigma^2 I]^{-1}(\boldsymbol{y} - m(T)) \tag{3.9}$$

The covariance matrix remains unchanged with the use of the prior mean. However, the use of a prior mean can introduce a strong bias towards the predictive mean and should be used with care. The second piece of information that can be utilised in *GP* regression is the noise of the data. Typically, the measurement noise is defined to be invariant for each time point, as a Gaussian distribution with a standard deviation σ. This assumption generates the noise term $\sigma^2 I$ present in the above equations. The established methodology for designing experiments of OD and fluorescence measurements dictates that both technical and biological replications of each different strain/condition must be generated. The technical replicates are present in the same plate in different wells with exact same conditions and inoculated from the same liquid culture, while the biological replicates are run on different occasions, starting from separate liquid cultures again in the same conditions. In this case, the noise term differs for each time point and may be calculated from the variance between the technical replicates. The only change to the calculations involving equations 3.4, 3.5 and 3.8 is that a diagonal matrix Σ populated by the variances of the different time points, replaces the term $\sigma^2 I$.

Figure 3.1 shows the resulting models of the fitting process using simple Gompertz fitting (3.1a), zero-mean *GP* fitting (3.1b) and Gompertz mean *GP* fitting (3.1c). In this instance, the *GP* with a Gompertz prior mean function outperforms the other models. The experimental data points are drawn from an instance

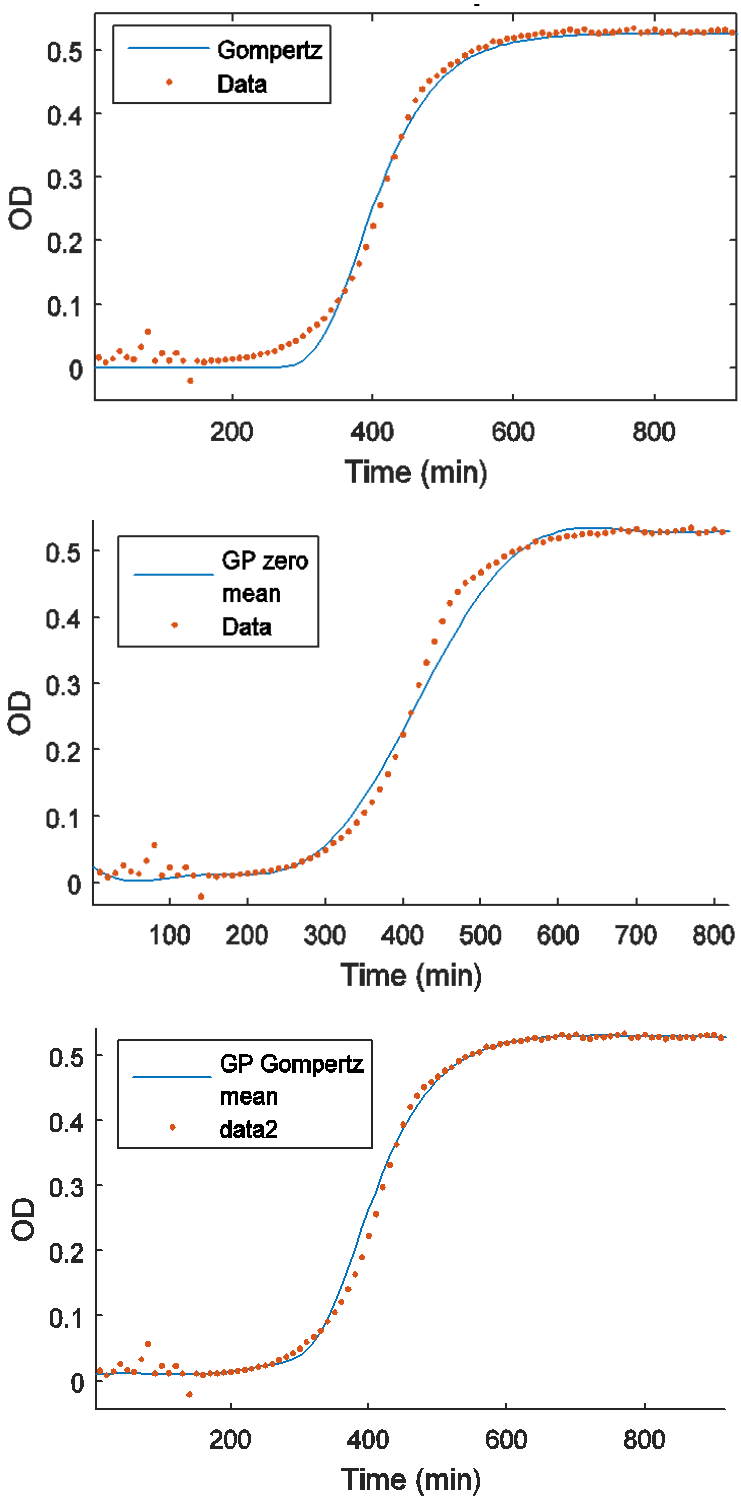


Figure 3.1: Fitted experimental data for OD with Gompertz model (top), *GP* with zero mean (middle) and *GP* with Gompertz prior mean (bottom). The combination of methods in the last case performs better.

of *E. coli* DH10B growth on M9 minimal media, with four technical replicates. Calculations were made with log-likelihood optimisation of the hyperparameters in both *GP* cases (3.1b and 3.1c) in accordance with an unmonitored, automated high-throughput process.

3.2 Mathematical modelling of promoter activities

This section will devote to the modelling approach that was adopted for the calculation of promoter activities per OD unit during the exponential phase of bacterial growth. The full algorithmic approach for the extraction of the exponential phase will be presented here and the *GP* approach will be extended for the fluorescent output as well. A presentation of the tool that was created for this purpose will follow.

The general purpose behind the study presented in this chapter is the definition of a new metric for the calculation of the activity of a promoter, mainly during the exponential phase of bacterial growth but not limited to it. The reason why most studies involving promoter characterization use only the exponential phase interval for the relevant measurements, is that it has been shown that during this phase, bacterial growth is optimum, therefore there are no external signals or factors limiting protein expression, contrary to the lag and stationary phases. However, using the same argument, the metric proposed here could be used in time intervals which include parts of the lag and stationary phases, where the role of particular regulatory elements tied to genes that are known to take part in the initiation and termination of the exponential phase, can be unravelled.

Recent research involving bacterial growth monitoring and indirect measurements

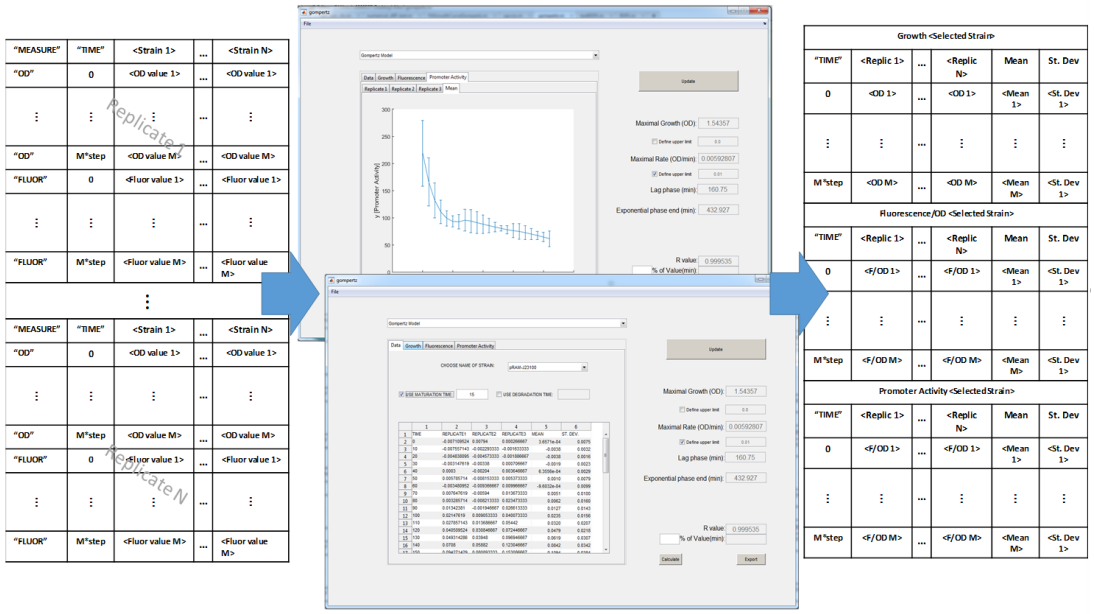


Figure 3.2: The input/output file formats and a general illustration of the tool.

of the activity of a gene, enlists devices known as plate readers. These can be either utilized independently or as a part of a larger high-throughput automated platform [72] and they combine a set of measurement types that can be performed independently or concurrently. The typical output of a plate reader constitutes a set of time-series replicated measurements of the OD and Fluorescence of bacterial cultures, growing in a microtiter plate consisting of a number of wells, corresponding to different strains and conditions. The reader output is usually represented in a data-sheet format as shown in figure 3.2.

Traditionally, the promoter activity for each time point has been expressed as the fluorescence intensity fluorescence intensity units (FU), in arbitrary units (FU) divided by the corresponding OD [73], the absorbance of light of a given wavelength over a defined optical path (dimensionless). The ratio FU/OD is a proxy for the amount of protein in mature state per cell. Other works have adopted a more broadly accurate approach by defining the total promoter activity as the difference in fluorescence between two distant time points divided by the integral of the function of growth [74]. However, this approach averages the

behaviour of the promoter over a substantial time course, rendering it suitable only for regulatory systems that maintain expression of the reporter gene at a near constant level (constitutive promoters). The new metric proposed in this chapter, is proven mathematically from previous models. This metric, while remaining simple in its form, captures the divergence of intensity values for two Green Fluorescent Protein (GFP) reporters of different stabilities, bound to the same promoter, providing evidence that it is a more accurate metric for the quantification of promoter activity.

3.2.1 Fluorescence time-series piecewise-polynomial regression

The implemented algorithm for the calculation of the promoter activity requires the definition of limits in the exponential phase which is based on the Gompertz model fitting (see section 3.1.1), the OD output and the FU output corresponding to the time interval of the exponential phase. The FU output is, like OD, a set of measurements corresponding to particular time points. These can be illustrated as scatter plot. The promoter activity as will be made evident below, requires a continuous function corresponding to growth, namely $G(t)$ which is the result of one of the fitting processes to OD data points, presented in 3.1.1 and 3.1.2, and a continuous function $I(t)$ corresponding to fluorescent output which must be twice differentiable as the promoter activity form contains both the first and second derivatives of fluorescence.

Let I_n, $n = 1, ..., N$ denote the N measurements of the fluorescence output of the liquid culture, corresponding to t_n, $n = 1, ..., N$ times as obtained from the plate reader. Two methods for the computation of $I(t)$ are implemented in this study. The first approach relies on the consideration of cubic polynomial functions as the basis for fitting the time-series (I_n, t_n). The best fit for fluorescence data is

calculated using a constrained free-knot polynomial approximation with a knot removal method illustrated in [75], [76]. The choice of cubic polynomials is argued on the fact that the resulting curve should be at least twice differentiable and the second derivative must contain information regarding the maturation process, as will be evident from the promoter activity formula (see equation 3.22). In effect, given the fluorescence data points, the algorithm's output for $I(t)$ is a piecewise polynomial of M pieces of the form:

$$I(t) = \begin{cases} a_1 t^3 + b_1 t^2 + c_1 t + d_1 & 0 < t \leq e_1 \\ a_2 t^3 + b_2 t^2 + c_2 t + d_2 & e_1 < t \leq e_2 \\ \vdots & \vdots \\ a_M t^3 + b_M t^2 + c_M t + d_M & e_{M-1} < t \leq e_M \end{cases} \tag{3.10}$$

The values of all the polynomial coefficients $(a_1, b_1, c_1, d_1), \ldots, (a_M, b_M, c_M, d_M)$, as well as the time intervals $0, e_1, e_2, \ldots, e_M$ and the number of polynomials M, are all determined by the algorithm.

3.2.2 Analytical formula for promoter activity

For the calculation of promoter activities, the computational tool presented in this work is based on previously described deterministic differential equations [77], but also taking into account the number of bacteria inside a liquid culture which changes as a function of time according to $G(t)$. Specifically, the model consists of a system of two first-order differential equations that describe the production of FP of the whole population of cells. The equations used contain the following elements: $P(t)$ represents the current 'non-mature' protein quantity per OD unit and $G(t)$ is the progression of bacterial growth with respect to time, measured in OD units. Therefore, $P(t)G(t)$ is the total non-mature protein contained in

the culture. Similarly, $P_m(t)$ denotes the 'mature' protein quantity per OD and $P_m(t)G(t)$ is the total mature protein in the culture. Then, according to the model:

$$\frac{d}{dt}[P(t)G(t)] = A(t) - k_m[P(t)G(t)] - k_d[P(t)G(t)] \tag{3.11}$$

$$\frac{d}{dt}[P_m(t)G(t)] = k_m[P(t)G(t)] - k_d[P_m(t)G(t)] \tag{3.12}$$

The values of k_m and k_d correspond to maturation and degradation rates, respectively. Both mature and non-mature versions of the FP are assumed to be susceptible to degradation. Both parameters (k_m and k_d) are obtainable from the values of maturation and degradation times - t_m and t_d - from the simple arithmetic $k_d = \frac{ln2}{t_d}, k_m = \frac{ln2}{t_m}$. The degradation time can be thought as the half-life of the protein and is assumed to be the same in both mature and non-mature states. The maturation time can be thought as a "sort of half-life" but with respect to the conversion of the protein from the non-mature to mature state. These values are characteristic to each different FP and must be either determined experimentally or obtained from literature.

The term $A(t)$ present in 3.11 signifies the current FP production rate performed by the entire bacterial population with respect to time. The promoter activity per OD then is defined as: $A_p(t) = A(t)/G(t)$. Equations 3.11 and 3.12 are then simplified by expanding the derivatives of the left-hand side, and dividing by G(t):

$$\frac{dP(t)}{dt} = A_p(t) - \frac{1}{G(t)}\frac{dG(t)}{dt}P(t) - k_mP(t) - k_dP(t) \tag{3.13}$$

$$\frac{dP_m(t)}{dt} = k_mP(t) - \frac{1}{G(t)}\frac{dG(t)}{dt}P_m(t) - k_dP_m(t) \tag{3.14}$$

The middle term of the RHS in 3.14 present in 3.13 as well, is also known as the 'dilution term' because it signifies the dilution of FP per cell, during cell growth. For a more intuitive comprehension of the latter, one can imagine a halt

of the FP production, while the population of the bacteria continues to increase. That would result in the dilution of already existing protein between the dividing cells. As mentioned previously, by using a plate reader it is possible to obtain OD and fluorescence values at defined time points, simultaneously. The continuous function fitted upon the time-series measurements of OD, denoted as G(t), and the continuous function corresponding to the fitting of the fluorescence measurements, defined as I(t), are both usually measured in arbitrary (dimensionless) units. Assuming that the measured fluorescent signal is proportional to the current mature protein quantity in the culture, the equation that describes the relation of the fluorescent signal with the mature protein per OD unit is:

$$I(t) = G(t)P_m(t) \tag{3.15}$$

Differentiating $I(t)$ and dividing by $G(t)$ yields:

$$\frac{1}{G(t)}\frac{dI(t)}{dt} = \frac{1}{G(t)}\frac{dG(t)}{dt}P_m(t) + \frac{dP_m(t)}{dt} \tag{3.16}$$

Solving with respect to the last term and substituting into (3.14), yields the following equality:

$$\frac{1}{G(t)}\frac{dI(t)}{dt} - \frac{1}{G(t)}\frac{dG(t)}{dt}P_m(t) = k_m P(t) - \frac{1}{G(t)}\frac{dG(t)}{dt}P_m(t) - k_d P_m(t) \tag{3.17}$$

The dilution term is simplified from the LHS and the RHS of the equation and $P_m(t)$ is substituted based on equation 3.15 :

$$\frac{1}{G(t)}\frac{dI(t)}{dt} = k_m P(t) - k_d \frac{I(t)}{G(t)} \tag{3.18}$$

The above equation when solved with respect to $P(t)$ yields expression which depends only on known outputs:

$$P(t) = \frac{1}{G(t)} \frac{(\frac{dI(t)}{dt}) + k_d I(t)}{k_m} \tag{3.19}$$

For simplicity of the following calculations, let $P(t) = Q(t)/G(t)$, with $Q(t)$ being the second term of the above product. Substituting $P(t)$ in (3.14) and expanding the derivative of $P(t)$ in the LHS yields:

$$\frac{1}{G(t)}\frac{dQ(t)}{dt} - \frac{1}{Gt)}\frac{dG(t)}{dt}\frac{Q(t)}{G(t)} = A_p(t) - \frac{1}{G(t)}\frac{dG(t)}{dt}\frac{Q(t)}{G(t)} - k_m\frac{Q(t)}{G(t)} - k_d\frac{Q(t)}{G(t)} \tag{3.20}$$

The derivative of $P(t)$ is rearranged in such away so that it is evident that the RHS second term is simplified with the LHS second term. Therefore:

$$\frac{1}{G(t)}\frac{dQ(t)}{dt} = A_p(t) - k_m\frac{Q(t)}{G(t)} - k_d\frac{Q(t)}{G(t)} \tag{3.21}$$

Then solving with respect to $A_p(t)$ and substituting $Q(t)$ finally yields:

$$A_p(t) = \frac{k_d(1 + (k_d/k_m))I(t) + (1 + 2(k_d/k_m))(dI(t)/dt) + (1/k_m)(d^2I(t)/dt^2)}{G(t)} \tag{3.22}$$

Equation 3.22 constitutes a general expression for the promoter activity as a

function of time which is based only on the functions G(t) and I(t) obtained by fitting the time-series data for OD and fluorescence, along with the degradation and maturation times. The tool uses the above formula to calculate the promoter activity at each time point based on user-defined values for the degradation and maturation times t_d and t_m. Using the same approach, simpler expressions can be obtained for the promoter activity for the cases, where one of maturation and degradation rates is omitted or both. When the maturation time of the FP is extremely rapid ($t_m < 10min$), which is usually the case, the maturation process is omitted and $P_m(t)$ is equal to $P(t)$. Therefore, 3.14 is not taken into account and the maturation term $k_m P(t)$, is omitted from 3.13. Solving 3.13 with respect to $A_p(t)$ then yields:

$$A_p(t) = \frac{k_d I(t) + (dI(t)/dt)}{G(t)} \tag{3.23}$$

When degradation is negligible, both 3.13 and 3.14 are used, but with the last terms corresponding to the degradation process omitted, pertaining to the case when extremely stable FPs are used ($t_d > 24h$). Using a similar numerical procedure, equation 3.22 changes to:

$$A_p(t) = \frac{(dI(t)/dt) + (1/k_m)(d^2I(t)/dt^2)}{G(t)} \tag{3.24}$$

Finally, the simplest case corresponds to when both maturation and degradation rates are omitted. This yields:

$$A_p(t) = \frac{(dI(t)/dt)}{G(t)} \tag{3.25}$$

The last and simplest case has been used in several studies, either as it is [78],

[79] or as an integrating function from which total promoter activity between two time points can be calculated [80].

The functionality of the tool is briefly shown in figure 3.2. The input is accepted in a comma-separated values (CSV) format as indicated, where the biological replicates are placed one below the other. The tool recognises different biological replicates for one single strain/condition and it dynamically creates separate tabs that show growth, fluorescence and promoter activity corresponding to the exponential phase. The 'update' button starts the procedure, calculating the promoter activity and FU/OD for the selected strain and plotting the graphs for each biological replicate and for the average of all the replicates. The promoter activity calculations are based, as mentioned before, on the values for maturation and degradation time provided in the user interface. There is also an 'export' functionality present in the tool, which outputs growth, fluorescence divided by OD and promoter activity data points for the selected strain.

3.2.3 Gaussian process regression of fluorescence

As illustrated before, the latest method which has been proven in [71] to be more accurate in the regression of biological relative data is *GP* regression. In particular, it has been shown, using synthetic data, that *GP*s with a radial basis (eq. 3.6) or a neural network covariance function outperform other regression methods such as local polynomial regression or smoothing spline. In the case of the second time-derivative of the underlying latent function which is estimated by the regression method, the *GP* based regression with a radial basis covariance function exhibits the higher accuracy. The latter observation constitutes the critical reason for our choice of this particular combination (*GP* with radial basis covariance function) for implementation in the tool, as the general promoter

activity formula (eq. 3.22) contains both first and second time derivatives of the function.

It has been shown in [81] that the derivatives of GPs are GPs themselves. Starting with a covariance function as described in equation 3.6, the first and second time derivatives are defined as follows:

$$\partial_1 k(t_i, t_j) = \frac{\partial k(t_i, t_j)}{\partial t_i} = -\frac{(t_i - t_j)}{l^2} k(t_i, t_j) \tag{3.26}$$

$$\partial_2 k(t_i, t_j) = \frac{\partial k(t_i, t_j)}{\partial t_j} = \frac{(t_i - t_j)}{l^2} k(t_i, t_j) = -\partial_1 k(t_i, t_j) \tag{3.27}$$

$$\partial_1^2 k(t_i, t_j) = -\frac{1}{l^2} k(t_i, t_j) + \frac{(t_i - t_j)^2}{l^4} k(t_i, t_j) \tag{3.28}$$

$$\partial_2^2 k(t_i, t_j) = \partial_1^2 k(t_i, t_j) \tag{3.29}$$

The predictive mean first and second derivative functions are therefore:

$$E[(f')^*] = \partial_1 K(T^*, T)[K(T, T) + \sigma^2 I]^{-1} \boldsymbol{y} \tag{3.30}$$

$$E[(f'')^*] = \partial_1^2 K(T^*, T)[K(T, T) + \sigma^2 I]^{-1} \boldsymbol{y} \tag{3.31}$$

It is evident that the derivatives only exhibits a single difference with respect to the mean predictive function 3.4, in effect the first factor of the RHS. Likewise, the covariance matrices for the first and second derivative functions are:

$$\boldsymbol{C}[(f')^*] = \partial_1\partial_2 K^* - \partial_1 K(T^*,T)[K+\sigma^2 I]^{-1}[\partial_1 K(T^*,T)]^T \tag{3.32}$$

$$\boldsymbol{C}[(f'')^*] = \partial_1^2\partial_2^2 K^* - \partial_1^2 K(T^*,T)[K+\sigma^2 I]^{-1}[\partial_1^2 K(T^*,T)]^T \tag{3.33}$$

where $K^* = K(T^*,T^*)$ and $K = K(T,T)$. The diagonals of the covariance matrices for the predictive mean, the first and the second derivatives (equations 3.5, 3.32, 3.33) indicate the variances in each time point for the predictions. The upper and lower bounds (UB and LB, respectively) of the 95% confidence interval can then be calculated as:

$$UB = E[f^*] + 2\sqrt{V(f^*)} \tag{3.34}$$

$$LB = E[f^*] - 2\sqrt{V(f^*)} \tag{3.35}$$

The vector $V(f^*)$ corresponds to the appropriate covariance matrix.

3.3 Molecular biology methods

This section deals with the molecular biology procedures that were used in the creation of the validation system for the modelling of growth curves and promoter activities. Plasmid creation methods, bacterial strains used, media, measurement equipment, FPs etc., are described here. Most of the work that is reported in this section was performed by Aitor de Las Heras.

3.3.1 Strains and growth conditions

The *E. coli* strains used for the experiments presented in this work are the DH10B strains carrying different promoter/reporter fusions. *E. coli* strains were grown in Lysogeny Broth (LB) base media or M9 minimal media supplemented with the antibiotic kanamycin at a concentration of 50 μg/ml, ampicillin at 50 μg/ml or chloramphenicol at 30 μg/ml to ensure plasmid maintenance. Strains were grown in 200 μl of medium at a temperature of 37^{o}C which has been proven to be optimal for the growth of *E. coli* and 700 rpm of rotary shaking was used. Standard molecular biology procedures were carried out according to Sambrook et al. [82].

Bacterial growth was estimated by measuring the OD of the cultures at a wavelength of 600 nm (OD600). OD600 and fluorescence values were measured a microplate reader FLUOstar Omega (BMG Labtech^{TM}). Bacterial cultures were propagated using 200 μl of LB/well. Measurements of fluorescence (lex 485 nm, lem 520 nm, gain 1300) and OD600 were obtained using black, 96-well microplates with flat, clear well-bottoms (Greiner). The strains containing the promoters pPtrc and pPm were induced with IPTG (SIGMA) or m-toluic acid (SIGMA) respectively, to a final concentration of 1 mM when they reached the mid-exponential phase (OD = 0.6).

3.3.2 Construction of plasmids

Plasmids carrying *yfp* gene controlled by different promoters were constructed as follows: a 749 bp fragment carrying the promoterless Yellow Fluorescent Protein (YFP) (venus) coding region was amplified by polymerase chain reaction (PCR) using as a template the pYZ2-123 [83] and primers AM3F/AM4R. The PCR product carrying the promoterless *yfp* coding gene downstream of

AM3F	TTCTGCAGTGAGACCGGACAATTAACAGTTAACAAATAAGGTCTCTATGAGCAAAGGTGAAGAACTG
AM4R	CCCAAGCTTTTATTTATACAGTTCGTCCATAC
J23Fw	CCCGGTCTCTGCAGCGCGGCCGCATCTAGAG
J23Rv	CCCGGTCTCATCATCTAGTATTTCTCCTCTTTCAC
LacIq1Fw	CCCGGTCTCTGAGTCAATTCAGGGTGGTGAATGT
LacIq2Rv	CCCGGTCTCATCATCTAGTATTTCTCCTCTTTCGTGTGAAATTGTTATCCGCTC
Ara1Fw	CCCGGTCTCTGCAGTCTAGGGCGGCGGATTTGT
XylS5R	CCCGGTCTCATCATCTAGTATTTCTCCTCTTTCATTGTTTCTGTTGCATAAAGCC

Table 3.1: Primers for the construction of the fluorescent protein expression plasmids with backbone pRAMv2.

two *Bsa*I sites and flanked by *Pst*I/*Hin*dIII restriction sites was cloned into pSEVA331 [84] generating pRAMv2. Two 92 bp fragments containing a strong (pJ23100) and a weak (pJ23114) constitutive promoter [85] were amplified using the primers J23Fw/J23Rv and as templates the Biobrick^{TM}parts BBa_J23100 and BBa_J23114, respectively. Another 1589 bp fragment carrying the lacIq coding gene and Ptrc promoter was amplified using LacIq1Fw/LacIq2Rv primers and pSEVA434 [84] as a template. Finally, a 1363 bp fragment containing XylS transcriptional factor coding gene and its cognate promoter Pm was amplified using the Ara1Fw/XylS5R primers and pSEVA328 as a template [84]. All primers are shown in table 3.1. These four fragments, which contained pJ23100 and pJ23114 constitutive promoters and LacIq/Ptrc and XylS/Pm inducible systems upstream of a strong Ribosome Binding Site (RBS) (BioBrick Part: BBa_B0034), were then cloned upstream of the *yfp* gene of pRAMv2 after being digested with BsaI generating pRAM-J23100, pRAM-J23114, pPtrc and pPm, respectively. All these plasmids also contain the T0 rho-independent transcriptional terminator of phage lambda (BioBrick Part: BBa_K864600) downstream of the *yfp* gene and were introduced in *E. coli* DH10B by chemical transformation [82].

To validate the ability of the model to accurately compute promoter activities using data obtained with FPs exhibiting different degradation rates, *E. coli* DH10B was transformed to include two additional plasmids: pJ23100-GFPm3 [72] and pJ23100-GFPlva [86] which carried the constitutive PJ23100 promoter upstream of the GFPm3 and GFPlva coding genes. Both plasmids carried a strong RBS (BioBrick Part: BBa_B0034) between the PJ23100 constitutive promoter (BiobrickTM part BBa_J23100) and the GFP coding genes. The structure of these plasmids is described in 3.6.

3.4 Experimental validation

This section is devoted in the presentation and discussion of the analysis of the experimental data that were obtained from the combined OD-fluorescence time-series experiments run on *E. coli* strains carrying plasmids with different kinds of promoters in combination with different fusions of FPs. The data that were obtained from the full experimental procedure described above, are presented here in the following figures. The new metric for the promoter activity is compared with the old and the advantages are highlighted.

3.4.1 Constitutive promoter activity

The promoter activity and FU/OD of two synthetic constitutive promoters was calculated. One strong (pJ23100) and one weak (pJ23114) [85]. For that purpose, *E. coli* strains carrying pRAM-J23100 and pRAM-J23114 were grown on LB and a combination of OD600/fluorescence was recorded every 10 min for 18 hours, yielding a time-series for OD and fluorescence. The resulting time-series data of three biological replicates were processed with an implementation of the method of the new metric in MATLAB^{TM}. Figure 3.3 - left shows FU/OD, while the

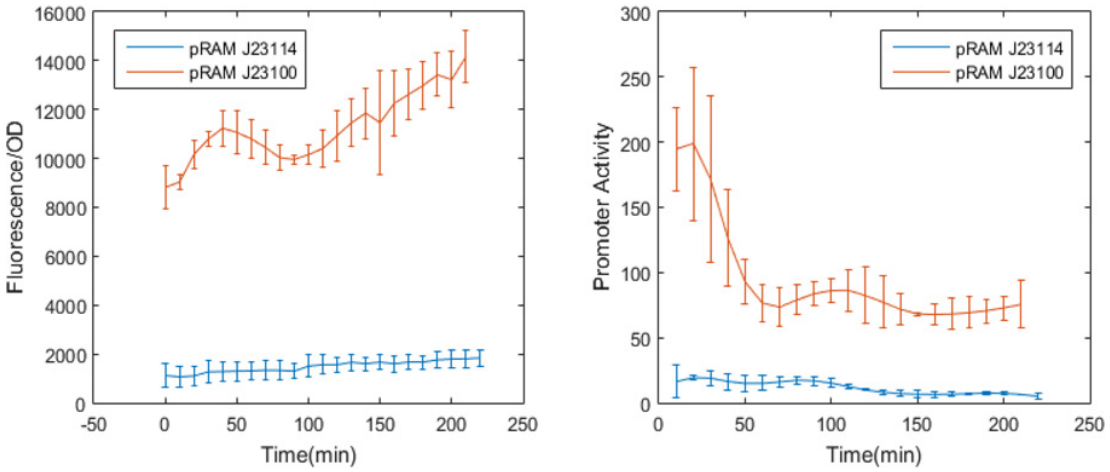


Figure 3.3: Constitutive promoter output analysis in terms of FU/OD (Left) and promoter activity (Right) displayed during the exponential phase of *E. coli* strains pJ23100 and pJ23114 growing in LB.

promoter activity as computed from the new formula is shown in 3.3 - right. Both graphs correspond to the exponential phase of growth which is extracted separately for each biological replicate as described in 3.1.1. The differences between promoter activity values of J23114 and J23100 promoters agree with the relative differences observed previous works (pJ23114 output has been reported as 10% of the pJ23100) [85], [87]. Moreover, the new metric outputs a function of time which describes the promoter activity perturbations during bacterial growth, rather than a single number that characterizes the overall strength. While the tool has been conceived to analyse the promoter activity during the exponential phase of bacterial growth, this approach can be used to identify changes in the activity of constitutive promoters at the end of exponential phase due to physiological control elements such as the replacement of the RNA polymerase subunits [88], [89] and the onset of other biologically important processes tied to the beginning or the end of the exponential phase. Comparing and contrasting the figures corresponding to the constitutive promoters, it is evident that the promoter activity in both cases starts with a relatively high value and steadily drops in an asymptotic manner to a lower value, following a trend that resembles damped oscillations.

3.4.2 Inducible promoter activity

Alternative methods for the characterization of the response of promoters have been reported [74], however, they reflect the behaviour of the promoters in a wide window of time and as such are suitable only for constitutive promoters, or for experiments that assume a constant activity during a particular time interval (e.g the activity of an inducible promoter after induction). In the case of inducible promoters or regulatory systems that exhibit a more complex translational activity, methods which do not capture the variation over time are not suitable for their characterisation. The ability of the new metric to illustrate the promoter activity over time, capturing the variations in a time-series presents a more suitable method capable of describing the behaviour of such complex elements.

For purposes of validation of the metric in this context, two different inducible regulatory systems were tested, namely LacIq/Ptrc [90] and XylS/Pm [91]. *E. coli* strains carrying pPtrc and pPm were grown in LB medium and OD600 and fluorescence measurements were taken every 10 min. Once they reached mid-exponential phase of growth, inducers (1 mM IPTG or m-toluic acid) were added to the medium. The results as shown in 3.4 correspond to the average of three biological replicates. The fluorescence and OD data of each replicate were processed using the implementation of the presented method with MATLAB for the calculatation of the FU/OD ratio (3.4 – left) and promoter activity as per equation 3.22 (3.4 – right). Both graphs correspond to the exponential phase of both strains, as extracted by the tool. The results indicated that the LacIq/Ptrc IPTG induction is stronger than the one mediated by m-toluic acid on XylS/Pm system. Additionally, FU/OD (Fig. 4 – left) data showed an increase of fluorescence on induction in both systems as a result of the accumulation of stable YFP inside the cells. Promoter activity revealed that both regulatory

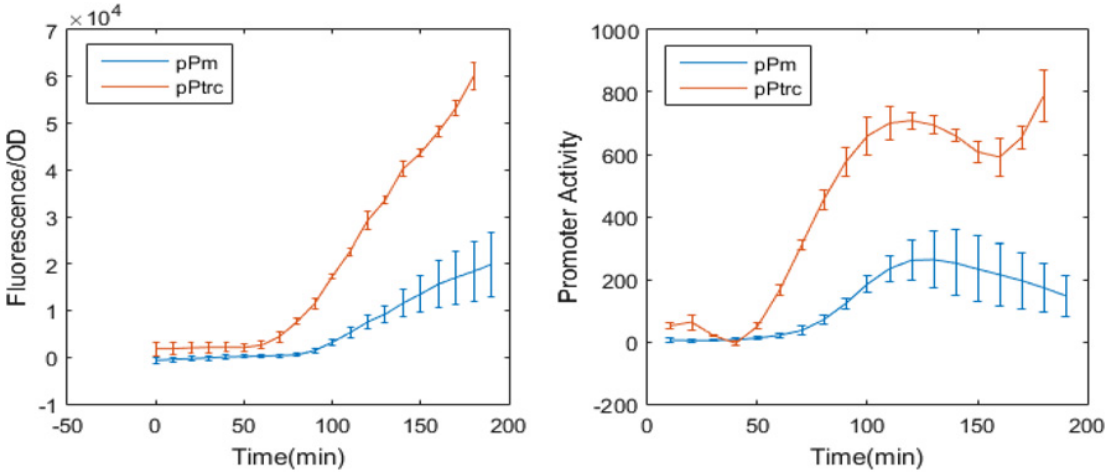


Figure 3.4: Inducible promoter analysis. (Left) FU/OD and (Right) promoter activities displayed during the exponential phase of *E. coli* strains bearing fusions of pPtrc and pPm with YFP, growing in LB. Induction of the system occurs at 50 minutes. The promoter activity metric yields more clearly pronounced and perhaps biologically relevant dynamics about the promoter.

nodes reached a plateau, which is in agreement with the known behaviour of these regulatory systems [92], [91].

3.4.3 Analysis of promoters using fluorescent proteins with different degradation rates

Most of the commonly used fluorescence proteins are extremely resistant to proteolysis [93] with a half-life of 24 h *in vivo* [94], [95]. This means that they will persist inside the cell even if their expression is shut down. Variants have been engineered which have significantly reduced half-lives [96], [97] with fluorescence shown to more closely follow the profile of mRNA abundance [77]. It follows that the degradation rates of fluorescent reporter proteins used for measuring promoter activities can have a significant impact on the final fluorescent output of a liquid culture. To illustrate this impact, an experiment was designed to include the same

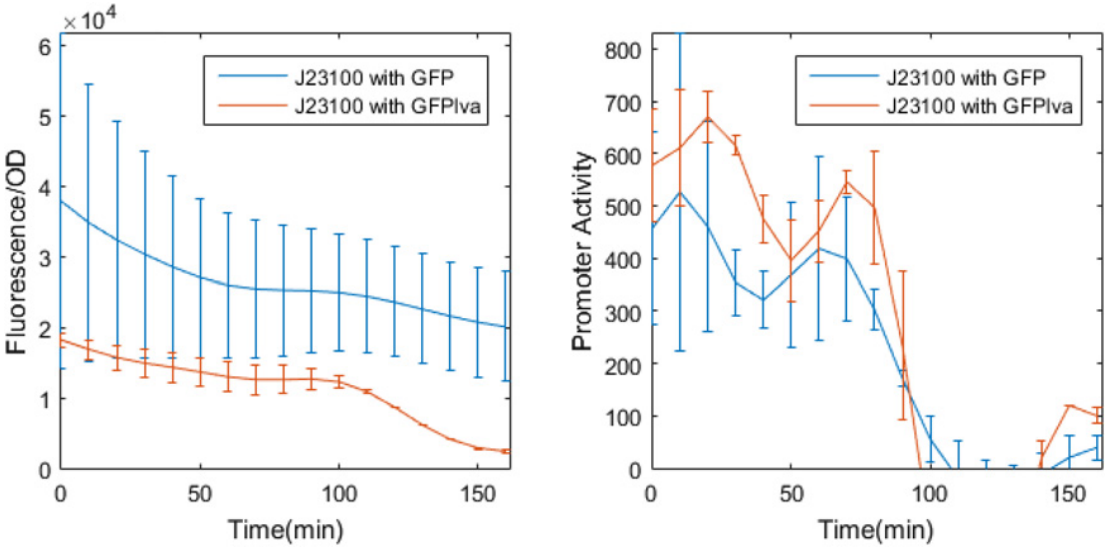


Figure 3.5: Promoter analysis using two versions of the GFP (left) FU/OD and (right) promoter activities displayed during the exponential phase of *E. coli* strains pJ23100-GFPm3 and pJ23100GFPlva growing in M9. The new metric comes closer to describing the same activity for the same promoter with different FPs.

promoter in the same strains coupled with a stable FP in one instance and an unstable in another. A way to quantify this impact is embedded in the analytical method applied for the calculation of promoter activities; the protein degradation rate is taken into account as shown in equation 3.22. To validate this approach with reporters of different half-lives, two *E. coli* strains were cultured in minimal media (M9), carrying the same strong constitutive promoter PJ23100 (BioBrick part: *BBa_J23100*) followed by a strong RBS (BioBrick Part: *BBa_B0034*) upstream of either a stable GFP (pJ23100-GFPm3) or a short half-life GFP (pJ23100-GFPlva). The experimental data consist of two biological replicates. Measurements of OD and fluorescence were taken every 10 min for 18 h. Results of this analysis are shown in 3.5. A comparison is made between the old metric (3.5 – Left) and the metric for promoter activities proposed herein (3.5 – Right), using the same set of data. Regression for the fitted curves was performed with *GP*s and the results shown correspond to the exponential phase of growth which

was defined as the interval of four doubling times before and after the maximum growth. The latter was calculated from the maximum value of the derivative of the growth curve ($\frac{d[G(t)]}{dt}$).

3.5 Interdependence of growth and fluorescence

The variation of bacterial growth curves will be explored in this section, the difference that it ensues in the final FP output will be illustrated. A case will be made again for the changing of metric based on this. This will include detailed mathematical analysis in Matlab^{TM}with figures.

3.5.1 Exponential phase duration

It has been made evident by now that the bacterial growth follows a sigmoidal trend in general, which is a macroscopic emergent behaviour in most biological system and a characteristic of population dynamics in a finite resources environment [98]. The Gompertz model is most times sufficient for the description of a typical bacterial growth, and as mentioned earlier, contains 3 parameters that fully define the model. However, bacterial growth is a process that can be influenced by many contributing factors in various aspects and even in such a simple growth model such as Gompertz, there are aspects of growth that are influenced by a combination of these parameters. One of these aspects worth looking into is the length of the exponential phase of growth. Assuming a Gompertz model for growth, the exponential phase duration can be calculated deterministically from the capacity (K) and the maximum growth rate (μ). According to equation (3.3) and the limits of the exponential phase set in section 3.1.1 (start: $t = \lambda$, end: $t_m, [\frac{dG(t)}{dt}]_{t_m} = \frac{\mu}{e}$) the length of the exponential phase is defined as $L_{exponential} = t_m - \lambda$. Assuming $\lambda = 0$ for simplicity of the calculations:

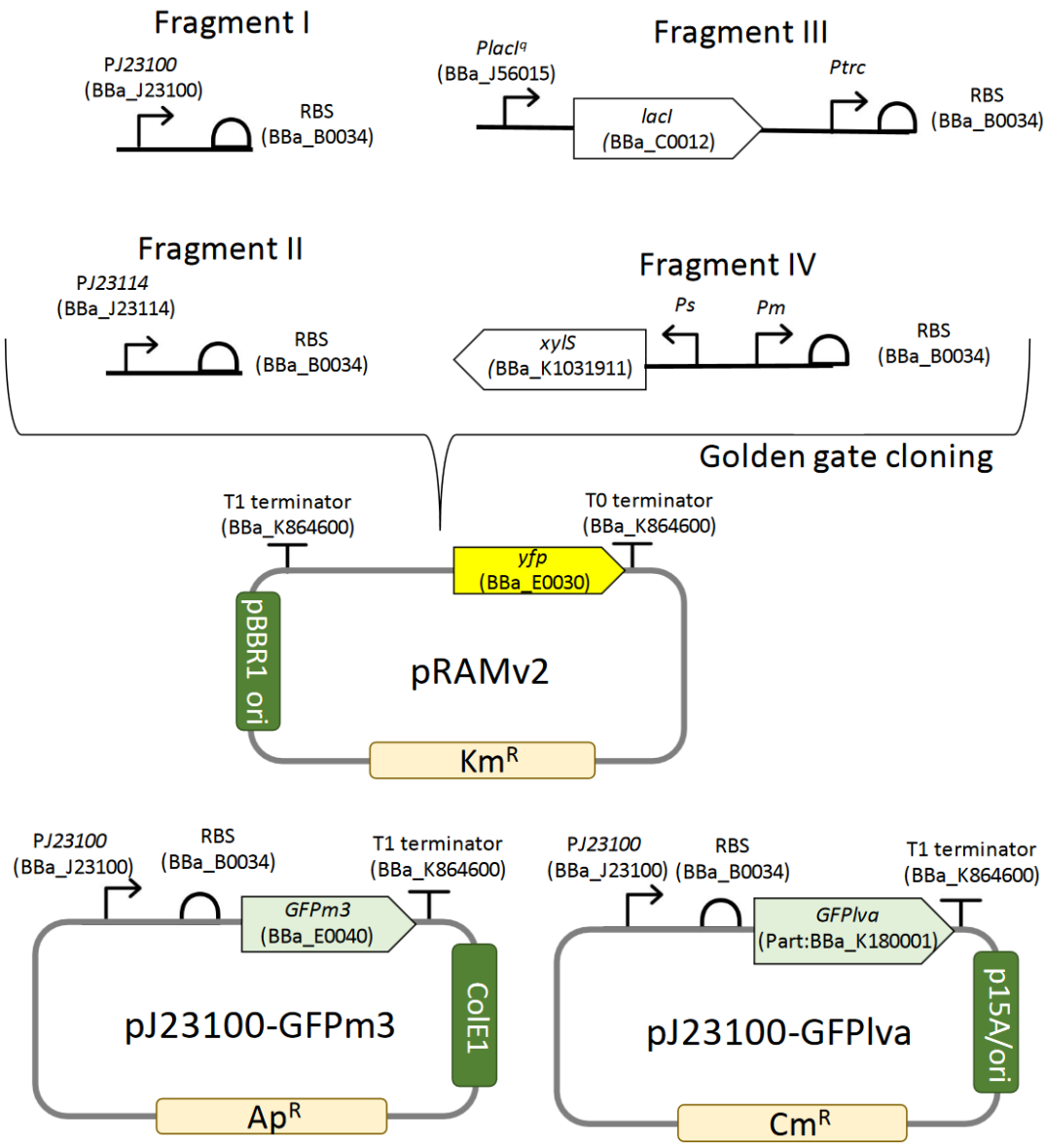


Figure 3.6: Schematic representation of the plasmids used in this study. Cloning strategy to generate pRAM-J23100, pRAM-J23114, pPrtc and pPm: fragments I, II, III and IV containing the pJ23100, pJ23114, LacIq/Ptrc and XylS/Pm regulatory elements and RBS (BBa_B0034) were assembled via Golden Gate reaction into pRAMv2 vector (Top). Schematic representation of the vectors pJ23100-GFPm3 and pJ23100-GFPlva which carry the pJ23100 promoter and the RBS (BBa_B0034) upstream of a stable (GFPm3) or an unstable GFP (GFPlva) (Bottom).

$$\mu \cdot exp[-exp[\frac{\mu \cdot e}{K} t_m + 1] + \frac{\mu \cdot e}{K} t_m + 2] = \frac{\mu}{e} \tag{3.36}$$

Solving the above equation will yield a solution for t_m which corresponds to the length of the exponential phase since we assumed that $\lambda = 0$. After simplifying the above with a few calculations:

$$e^{-x+1} + x - 3 = 0 \tag{3.37}$$

The x above is defined as $x = \frac{\mu \cdot e}{K} t_m$. The solution to this equation is given with the help of the lambert function, where only the real positive solution is considered as the only one with physical meaning in our case:

$$x = W(-\frac{1}{e^2}) + 3 = 2.841 \tag{3.38}$$

Therefore, substituting x from above:

$$L_{exponential} = t_m = 2.841 \cdot \frac{K}{\mu \cdot e} = 1.045 \cdot \frac{K}{\mu} \tag{3.39}$$

In essence, from the above calculations it is proven that the length of the exponential phase is proportional to the medium capacity (K) and inversely proportional to the maximum growth rate with a multiplying factor which is almost equal to one. This result is expected as the maximum growth rate determines the average growth rate during the exponential phase, which in turn influences the time needed for bacteria to reach their maximum density. On the other hand, the medium capacity corresponds to the maximum density of the bacteria, therefore a higher value is impacts the time needed to reach that value.

3.5.2 Fluorescent output during exponential phase and promoter activity

The above result shows that the exponential phase interval can vary with respect to the parameters of the model, which themselves are influenced by many changing factors during growth, such as temperature, richness in carbon sources, expression of particular proteins and many other exogenous or endogenous factors. Considering the ODE model described by equations 3.11 - 3.14 and assuming that constitutive promoters such as J23100 or J23114, maintain a constant level of activity during the exponential phase and a switch-like increase of the activity changing from lag-phase to exponential phase as well as a similarly abrupt decrease during transition from the exponential phase to the stationary phase, it will be made evident that the length of the exponential phase plays a pivotal role in the final fluorescent output. More specifically, consider the assumption that $A_p(t)$ follows a clade function as follows:

$$A_p(t) = \begin{cases} \frac{a}{100} & t \leq \lambda - 10 \\ \frac{a}{100} + \frac{99a}{100}(3(\frac{t-\lambda+10}{10})^2 - 2(\frac{t-\lambda+10}{10})^3) & \lambda - 10 < t \leq \lambda \\ a & \lambda < t \leq t_m \\ a - a(3(\frac{t-t_m}{10})^2 - 2(\frac{t-t_m}{10})^3) & t_m < t \leq t_m + 10 \\ 0 & t_m + 10 < t \end{cases} \tag{3.40}$$

The transition in promoter activities between different phases is mediated with a smooth step function using Hermite polynomials in order for the clade function to remain differentiable across all time points. Let $a = 75\frac{FU}{OD \cdot min}$ as a representative activity for a strong constitutive promoter which is taken from the average value from the measurements of the J23100 promoter illustrated in 3.5. For purposes of

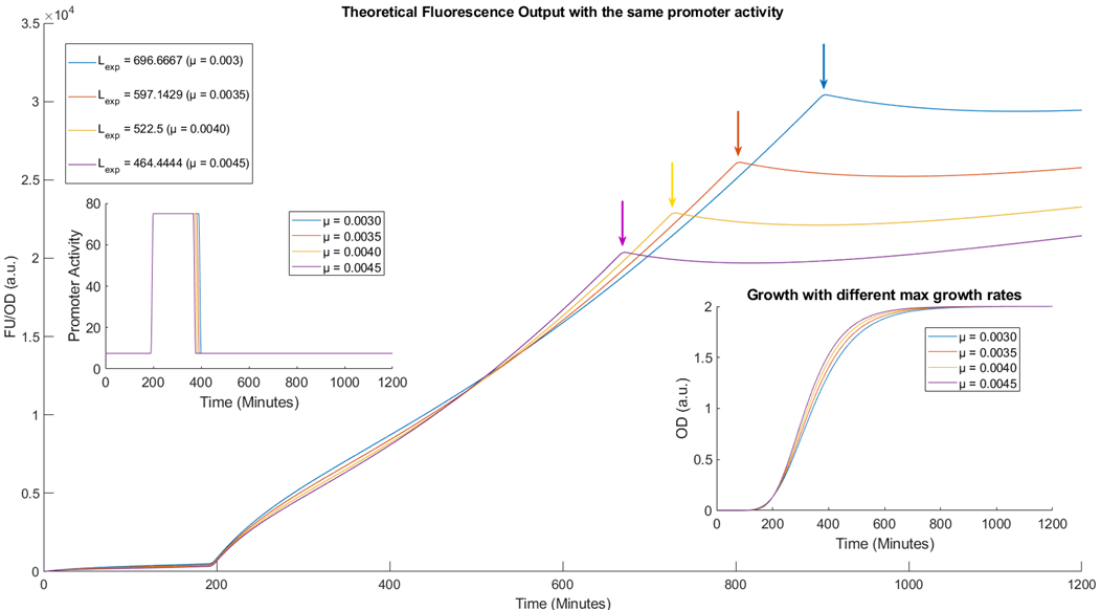


Figure 3.7: Fluorescence output between strains with the same promoter activity but different exponential growth windows. Fluorescence output is significantly higher at the end of the exponential phase which may result in misinterpreting the promoter activity.

simplicity, the maturation and degradation processes of the FPs can be omitted. Therefore the ODE model is reduced to:

$$\frac{dI(t)}{dt} = A_p(t)G(t) \tag{3.41}$$

Equation 3.41 was obtained from 3.11, 3.15 and the fact that because the maturation process is omitted, $P_m(t) = P(t)$, therefore, $I(t) = P(t)G(t)$ and $A(t) = A_p(t)G(t)$.

By integrating equation 3.41 and dividing by OD, the final normalized fluorescent output as per the old metric is given. As shown in chapter 5, bacterial populations even from the same species and minor differences in their genotypes are not consistent in their final OD (K) or their max growth rate (μ). Promoter activities

themselves can of course be influenced by changes in genotypes but even in the case that they are not, figure 3.7 shows that the old metric may misinterpret promoter activities when the maximum growth rates vary. The same is true for varying final OD (K) as they both influence the exponential growth window length.

Chapter 4

Strain development for *in vivo* fluorination and directed evolution

The main purpose of the work described in this chapter is the creation of a strain which can be utilized as a host for directed evolution experiments of the fluorinase enzyme. Choice of the base host and modifications to be implemented are based on an two - objective goal. One of the objectives is choice/construction of a suitable host for the genetic selection of high - performing variants of the fluorinase. The second objective is the creation of an *in vivo* bioreactor for fluorinated products. The host of choice is an *E. coli* variant, namely BL21(DE3), which has been extensively utilized for protein overexpression but also for the creation of novel bioreactors consisting of recombinant protein expression and other rationally designed alterations of the base host. Modifications performed are:

- The overexpression of a codon-optimized, recombinant version of the highest performing fluorinase enzyme known at the time of implementation.

- The deletion of a fluoride efflux pump gene (*crcB*) which has been shown to infer resistance to fluoride by expelling fluoride anions out of the cell.
- The deletion of the PNP gene (*deoD*), which catalyses the degradation of adenosine to sugar and phosphate.
- The expression of a SAM transporter in *E. coli*, a membrane channel protein which has been shown to actively transport SAM molecules from the extracellular medium inside the cell.

4.1 Fluorinase expression in *E. coli*

This section will include the work that has been done, resulting in the overexpression of a recombinant codon-optimized fluorinase in *E. coli* BL21(DE3). The *in vivo* version of the protein and the creation of the appropriate constructs for this will be described here as well, including plasmids, primers used, promoters, strong RBS, terminators etc.

4.1.1 Synthesis of the fluorinase gene

The first step towards expression of the fluorinase in *E. coli* BL21(DE3) is the synthesis of the fluorinase gene corresponding to the highest activity fluorinase as depicted in the study of O'Hagan' s group [53]. The most common approach for the expression of recombinant proteins in foreign hosts, includes designing appropriate forward and reverse primers and amplifying the desired gene with PCR followed by cloning of the gene to an appropriate expression plasmid. Recent advances in biotechnology however, enable the ordering of DNA sequences in relatively low prices. The gene synthesis along with the recipient plasmid were ordered from the GeneArt^{TM} service offered from Thermo Fisher Scientific. This

service allows for the ordering of synthesized genes accommodated in commercial constructs (plasmids) that include several features for optimized overexpression and convenient purification of recombinant proteins. The plasmid chosen for the accommodation of the fluorinase gene (*flA*1) is pET151/D-TOPO®. This is a low-copy expression plasmid, with a pBR322 origin of replication controlled by a ROP protein, and is maintained in the strain by ampicillin resistance selection. When growing the strains in the antibiotic ampicillin, the presence of the plasmid inside the strain is ensured. This particular plasmid contains the IPTG - inducible T7 promoter followed by the lac operator, upstream of the synthesized gene. This fusion is functional only in *E. coli* strains containing the T7 RNA polymerase which is naturally present in T7 bacteriophage and binds selectively on the T7 promoter to initiate transcription. The T7 promoter' s inducible nature along with its high activity, illustrated in the previous chapter, justifies its choice for overexpression - and optionally, subsequent purification - experiments. The lac operator is bound by a lac repressor to prevent basal expression in the uninduced state. The lac repressor, which is expressed constitutively from the plasmid, is silencing the expression of the fluorinase by binding both to the T7-lac promoter upstream of the fluorinase gene, preventing the T7 polymerase from attaching to the T7 promoter, and by binding to the lacUV5 promoter upstream of the T7 polymerase gene in the BL21(DE3) chromosome, preventing the expression of T7 polymerase altogether. However, a minute amount of basal expression still escapes repression as shown in 3.4. The fluorinase gene was synthesized with a codon-optimization strategy specific for *E. coli*. This is an established method for alleviating the bottleneck in expression rate of recombinant proteins, as the original sequence commonly contains codons that have different availabilities in the native host, which is the case for the fluorinase as well, whose native host is *Streptomyces sp.* MA37.

The synthesized gene is fused upstream with a 6xhistidine tag, which enables the optional purification of the protein by binding of the polyhistidine to nickel beads.

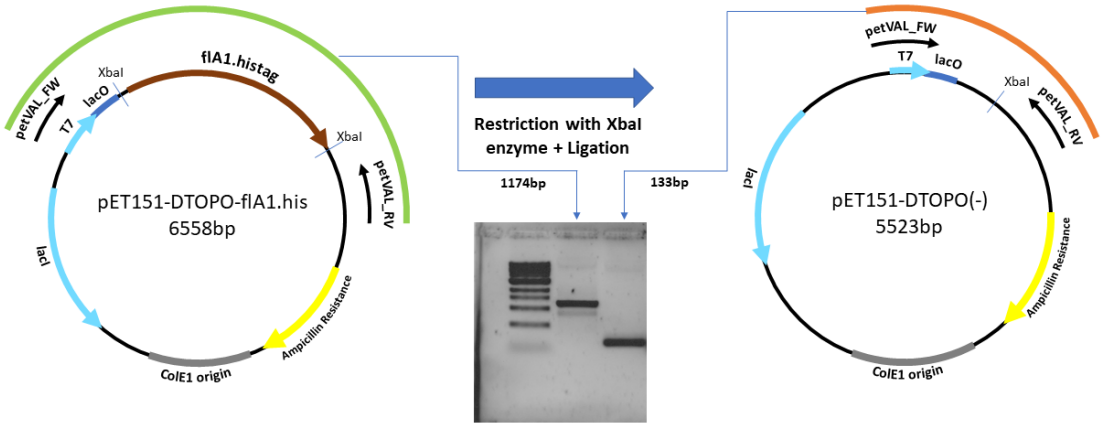


Figure 4.1: Plasmid pET151/D-TOPO including the codon optimized histidine-tagged *flA*1 gene (Left). Plasmid pET151(-) created with the excision of the *flA*1 gene by restriction from *Xba*I, serving as negative control (Right). Validation PCR showing the expected sizes (Middle).

A V5 epitope site follows, which can be used for recognition of the protein by the anti-V5 antibodies for western blot analysis. Finally, directly downstream of 6xHis and V5, a TEV recognition site resides for the removal of the 6xHis and the V5 epitope, by TEV protease treatment, a necessary step in order to obtain the original protein in purified form. Usually, the gene is placed between two different restriction sites as a result of the cloning process, in this case an *Xba*I and a *Sac*I site. A custom change was also included, specifically the addition of another *Xba*I site directly downstream of the sequence which was used to facilitate the one-step creation of the negative control plasmid. The final construct is illustrated in figure 4.1.

4.1.2 Cloning and transformation of strains

The plasmid pET151/D-TOPO accomodating the fluorinase as described above (hereafter referred as p*flA*1-his), was received in a glycerol stock tube containing the wild-type *E. coli* K12 with the plasmid. A sterile loop was used to top-off a small quantity and restrict the bacteria in an LB-agar plate with added ampicillin up to a final concentration of 50 μg/ml. The plate was left in 37 ^{o}C overnight and the next day single colonies where picked and transferred to falcon tubes containing 10ml of LB with ampicillin as previously, and again left to grow at 37 ^{o}C overnight. The liquid culture was subsequently centrifuged in 6000 g and the pellets were used for plasmid extraction, as described in section 2.5.3 in order to obtain the p*flA*1-his plasmid. The plasmid was then subjected to a restriction reaction with the enzyme *Xba*I. The reaction was left to incubate at room temperature for 1 hour and then it was loaded for gel electrophoresis. The uncut plasmid was also loaded as a control and the results are shown in figure 4.2. The bands that appear in the lane where the restriction reaction was loaded are in agreement with the sizes that are expected after the restriction, with the high molecular weight corresponding to the cut pET151/D-TOPO construct.

The next step was the creation of a negative control plasmid, in effect, the same backbone but with the fluorinase gene removed. To that end, the higher molecular weight band was cut from the gel with a scalpel, and then purified. The resulting pure DNA corresponded to a linearised plasmid restricted with *Xba*I and as a result, had compatible *Xba*I sticky ends. The latter enabled creation of the negative control plasmid with a ligation reaction along with an appropriate buffer and the purified cut plasmid at 5 μg/ml. The reaction was left to incubate in room temperature for 2 hours, resulting in the connection of the *Xba*I sticky ends to produce circular DNA. This one-step creation of the negative control was made possible by including an additional *Xba*I site at the end of the synthesized fluorinase gene. A chemical transformation process was then

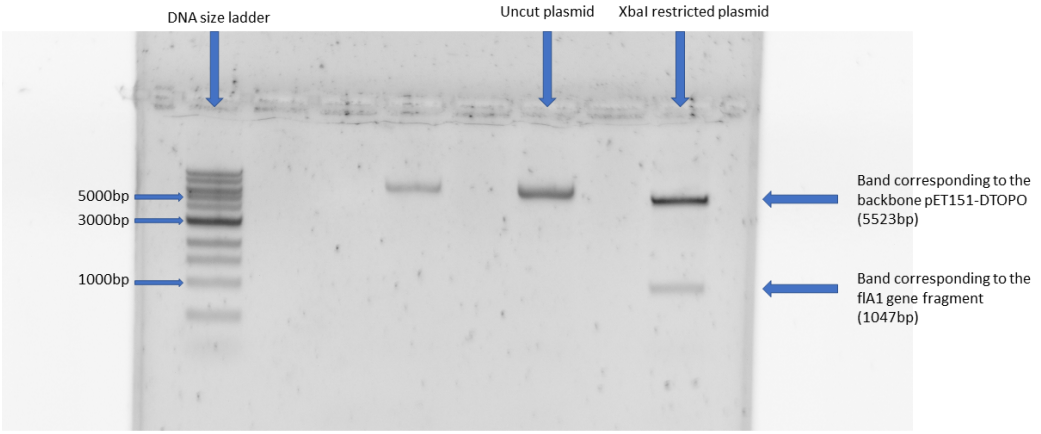


Figure 4.2: The pET151/D-TOPO with the histidine tagged *flA*1 gene as extracted from the strain *E. coli* K12 and the respective restriction with *Xba*I. The bands appearing in the restriction correspond to the expected sizes.

performed, using *E. coli* DH10B chemical competent cells. The transformation plates containing LB-agar with 50 μg/ml of ampicillin were left in 37 ^{o}C overnight and the next day, single colonies were picked for colony PCR. The negative control genotype was validated by using previously designed primers pet151VAL_FW and pet151VAL_RV (see table 4.1), upstream and downstream of the gene as shown in figure 4.2. The high molecular weight band (lane 1) corresponds to the original plasmid and the low weight band (lane 2) to the negative control, as expected for the amplified DNA lengths of the plasmid with and without the fluorinase gene.

Following the creation of the negative control (hereafter pET151(-)), the transformation of *E. coli* BL21(DE3) with both plasmids was performed. The same chemical transformation protocol as above was used and colonies from the resulting agar plates were tested with PCR to confirm the successful uptake of both

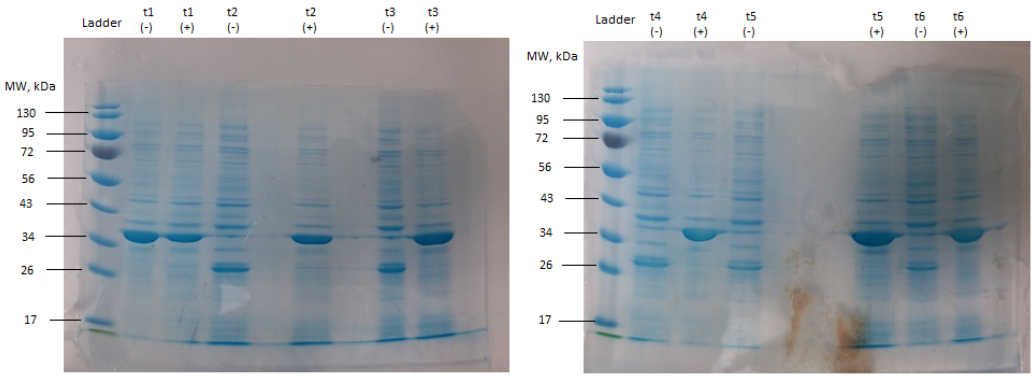


Figure 4.3: SDS page gel electrophoresis of the *E. coli* BL21(DE3) pilot expression of the fluorinase gene under the control of a lacI inducible promoter. Each band corresponds to samples obtained every one hour and is paired with the negative control plasmid pET151(-) grown in the same conditions. Optimal expression occurs 5 hours after induction (t5).

p*flA*1-his and pET151(-) plasmids as before with the expected amplified product weights.

4.1.3 Validation of fluorinase expression

As described in 2.1.1 *E. coli* BL21(DE3) is the host of choice for recombinant protein expression due to modifications that enable overexpression of proteins by the inducible high-strength T7 promoter. Both strains that were created from the above transformation were cultured in the same conditions as follows:

Single colonies of BL21(DE3) with p*flA*1-his and pET151(-) were put separately in two liquid LB tubes supplemented with ampicillin and left to grow overnight in 37 ^{o}C with 170 RPM shaking to saturation. The next day, two 250 ml flasks

pet151VAL_FW	TAATACGACTCACTATAGG
pet151VAL_RV	GCTAGTTATTGCTCAGCGG
FLA1_C_VV_FW	GG-ACTAGT-GAAAGAGGAGAAATACTAG-ATGGCAGCAAATGGTAGCCA
FLA1_C_VV_RV	CG-GGATCC-ACACTAGCACTATCAGCGTTA-TTAACGTGCTTCAACACGAAC
B0015_TER_FW	CG-GGATCC-CTACTAGAGCCAGGCATCAA
B0015_TER_RV	AA-CTGCAG-CGGCCGCTTCTAGTATATAAAC
FLA1_I_VV_FW	GC-TCTAGA-AATAATTTTGTTTAACTTTAAGAAGGAGATATACAT-ATGGCAGCAAATGGTAGCCA
FLA1_I_VV_RV	CG-GAGCTC-GCCCTT-TTAACGTGCTTCAACACGAAC
FLA1_INSEQ_FW	TATATTGCACCGAATAATGGTC
FLA1_INSEQ_RV	TTTCACGGCTATAAAAGGTCG
PJ61002_SEQ_FW	TCGCGGCCGCATCTAGAG
PJ61002_SEQ_RV	AGTCAGTGAGCGAGGAAGC

Table 4.1: Primers for the construction and validation of the fluorinase expression plasmids with backbones pET151/D-TOPO and pJ61002. The red colour signifies restriction sites, the blue colour corresponds to sequences inside pJ61002. The brown colour shows sequences that are part of the codon optimized *flA*1 gene and otherwise are the pET151/D-TOPO backbone residing sequences.

containing 50 ml of LB and ampicillin were inoculated with 500 µl (1:100) from the previous cultures and left to grow in 37 ^{o}C with 170 RPM shaking until OD reached a value of 0.7 - 0.8 (approximately 2 hours), which corresponds to the mid-exponential phase of growth. At that point, IPTG was added (in both cultures) to a final concentration of 0.5 mM to induce protein expression. The cultures were left to grow up to 6 more hours. Every 1 hour the OD of both cultures

Time (hours after induction)	Optical Density	
	p*flA*1-his	pET151(-)
0	0.809	0.684
1	2.149	1.562
2	2.905	1.822
3	3.147	2.021
4	3.352	2.123
5	3.131	2.240
6	3.349	2.359

Table 4.2: *E. coli* BL21(DE3) with p*flA*1-his and pET151(-). The OD values obtained were used for subsequent dilution of the samples to an OD of 1. The interesting effect of 30% higher growth of the strain expressing the fluorinase is also observed.

was measured (table 4.2) and an additional sample was taken from each culture, centrifuged in 20,000g for 30 seconds and the pellet was stored in -20 ^{o}C for future validation of fluorinase expression. Based on the OD values appropriate aliquots were taken and diluted with LB up to 500μl and a final OD of 1 to ensure equal culture density. The pellets gathered from the above procedure were analysed with SDS-page gel electrophoresis as follows:

Pellets were resuspended in 80 μl of 1xSDS-page sample buffer, put in a thermoblock at 100 ^{o}C in order to boil for 5 minutes and subsequently left to equilibrate in room temperature. For each of the time points 10 μl of both samples were loaded in a hand-crafted SDS-page gel. The gel was left to run for 1 hour and then subjected to Coomassie blue straining. More specifically, it was submerged to the stain and left to incubate for 1 hour to ensure sufficient staining. The resulting bands are shown in figure 4.3.

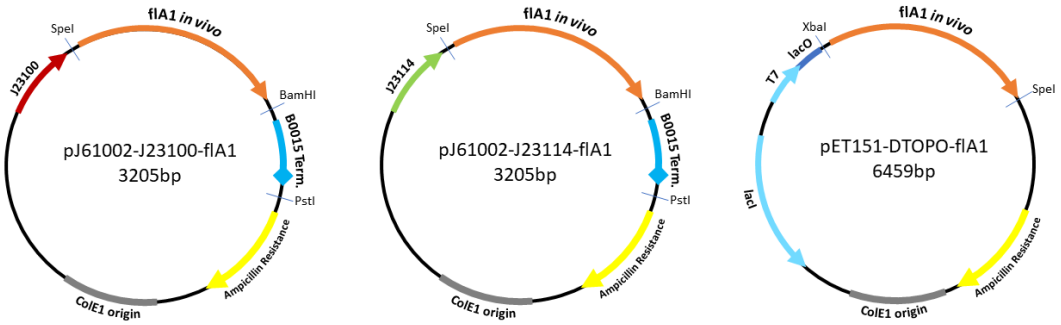


Figure 4.4: Three plasmids facilitating the *in vivo* version of the fluorinase gene: pJ61002-J23100-*flA*1 (left), pJ61002-J23114-*flA*1 (middle), p*flA*1 (right).

Gel lanes correspond to 6 time points and the samples with p*flA*1-his plasmid are placed side by side with the pET151(-) negative control. The band of interest appears in 34 KDa with a clear difference in intensity in the p*flA*1-his plasmid compared to the negative control, with the exception of lane 1, probably due to sample cross-contamination. Therefore, the expression of the enzyme was validated and the time point showing the highest expression is 5 hours after induction. This is considered the pilot experiment for expression, which not only confirms the expression of the recombinant protein but also pinpoints the optimal duration for purification or for initiating the *in vivo* reaction.

4.1.4 Alternative strains for fluorinase expression

As mentioned above, the synthesized gene was fused upstream with a polyhistidine tag, a V5 epitope and a TEV recognition site. It has been shown that in certain occasions enzymic activity is hindered when a polyhistidine tag is present in the amino acid chain, as reported in [99].

Therefore, to avert the possibility of impairment in the function of the fluorinase, it was deemed necessary to create complementary constructs of the fluorinase

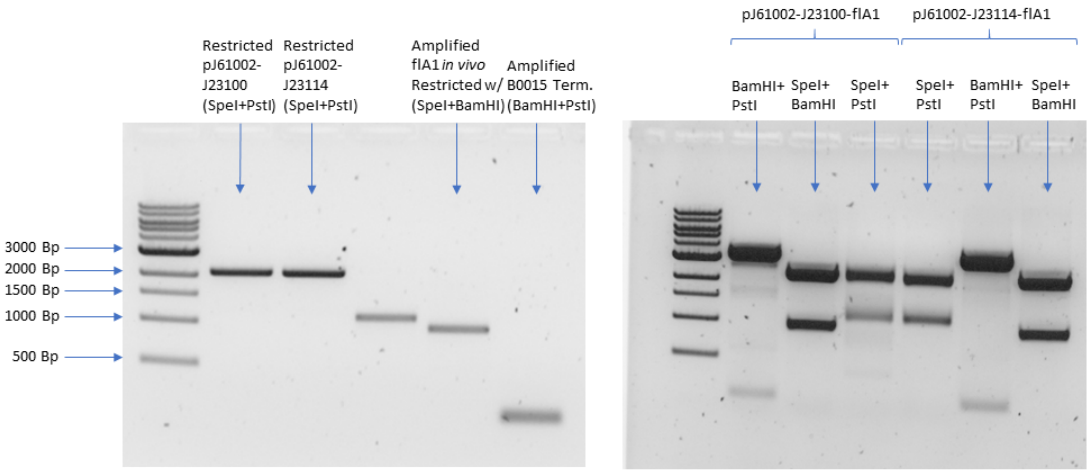


Figure 4.5: Left: Purified restriction of plasmid backbones pJ61002 bearing J23100 and J23114. Purified and restricted PCR amplification of the *flA*1 gene without his-tag and the B0015 terminator with compatible cohesive ends. Right: Confirmatory gels of the constructed pJ61002-J23100-*flA*1 and pJ61002-J23114-*flA*1 with appropriate restriction reactions.

gene without these fusions (*in vivo* version). To that end, three more plasmids have been created as shown in figure 4.4. In two of the plasmids, the pJ61002 backbone has been used to accomodate the *in vivo* version of the fluorinase.). One plasmid harbours a strong constitutive promoter (J23100) and the other accommodates a weak constitutive promoter (J23114). These plasmids were initially accommodating the MCherry Red Fluorescent Protein (RFP) protein used to quantify the promoter activities (see chapter 3). Both versions feature a strong RBS site and a B0015 strong terminator downstream. The lack of an appropriate restriction site between the terminator site and MCherry required an alteration in the strategy for the creation of the *in vivo* fluorinase constructs. Two sets of primers have been designed, namely FLA1_C_VV_FW, FLA1_C_VV_RV, B0015_TER_FW and B0015_TER_RV (table 4.1) and the corresponding products have been amplified with PCR using p*flA*1-his

and pJ61002 as templates. The first product (first two primers) is the *in vivo* fluorinase fused with *Spe*I and *Bam*HI restriction sites and the second is the *Bam*HI-*Pst*I fused B0015 terminator. Following purification of the PCR products, two restriction reactions were set-up with *Spe*I-*Bam*HI and *BamH*I-*Pst*I enzymes respectively in order to create compatible cohesive ends for the ligation reaction. The two pJ61002 plasmids were also added in two additional restriction reactions with enzymes *Spe*I and *Pst*I, in order to cut out the *mCherry* gene and the two linearised plasmid products were subsequently separated with gel electrophoresis. The bands corresponding to the linearised pJ61002-J23100 and pJ61002-J23114 were then excised from the gel and purified. In figure 4.5, the respective amplified products in an agarose gel after restriction as well as the restricted pJ61002 plasmids after purification. These 2 restricted linear DNA parts were combined with the appropriate restricted plasmids in ligation reactions to yield pJ61002-J23100-*flA*1 and pJ61002-J231114-*flA*1 respectively. The third *in vivo* fluorinase expression plasmid uses the same pET151/D-TOPO backbone but without downstream *Xba*I site and the *flA*1 gene lacking the his-tag (p*flA*1). All three plasmids accommodating the *in vivo flA*1 where sequenced with two additional primers (FLA1_INSEQ_FW, FLA1_INSEQ_RV) both residing inside the *flA*1 sequence. These were also used in validation PCR to ensure the correct alignment of the gene.

4.2 CrcB channel and fluoride availability

This section will describe the CrcB channel and display the reasoning for deletion of the gene corresponding to the CrcB channel protein. The procedure for deletion will be presented here in a detailed manner.

4.2.1 Inhibition mechanisms of fluoride

Fluoride has been shown to induce growth inhibition in bacteria, hence the extensive utilization of fluoride-based products for tooth care [100]. The mechanisms of inhibitions have been studied and include acidogenicity and acidurance [101]. Fluoride anions passively enter the bacterial membrane in the form of HF which is created from the free protons H^+ present in a water solution and free fluoride anions where it is again dissociated yielding free fluoride anions [102]. The presence of fluoride intracellularly within a range of molarities of the order of mM, acts in certain enzymes that are essential in the metabolism of bacteria and inhibits their activities. Two of these enzymes are F-ATPase and Enolase. F-ATPase is a membrane protein which directly controls the acidity of the intracellular environment by pumping out H^+ ions as part of the ATP hydrolysis action. Fluoride anions inhibit the activity while free H^+ cations acidify the intracellular envinronment. Enolase inhibition from fluoride is mediated by the formation of an ATP-like complex from ADP, free fluoride anions and divalent metal ions [103]. As a result, glucose metabolism is downregulated directly by the enolase participation in the pathway and indirectly because it produces phosphoenolpyruvate (PEP) which is essential for the PEP-dependent phosphotransferase system (PTS) by which glucose is uptaken from the environment. Another enzyme with a possible contribution in fluoride sensitivity is the pyruvate kinase, which is placed in the glucolytic pathway immediately after enolase, again directly influencing glucose metabolism.

4.2.2 Fluoride resistance in bacteria and eukaryotes

Because of the general nature of inhibition mechanisms of fluoride, it's toxicity is evident across all types of organisms. However, the fact that a range of organisms are naturally tolerant to high concentrations of fluoride, has driven research

efforts towards the identification of fluoride toxicity mitigation mechanisms that those organisms facilitate. Until recently, little was known about the putative mechanisms of actions or the genes associated with high fluoride concentration tolerance. Three recent studies revealed a wide variety of proteins that play a key role in dealing with a high fluoride concentration challenge in bacterial or eukaryotic hosts [104], [9], [105]. The identification of these proteins was made possible by studying fluoride specific riboswitches and identifying the genes whose expression they were controlling. Although they are distant in terms of phylogeny, the proteins that are produced in response to fluoride bear functional resemblance to CLCs (chloride channels). Protein conserved domain evidence initially misled to assigning these proteins as anion channels specific to fluoride but in contrast to other CLCs they are F^-/H^+ antiporters. In essence, these types of proteins are transmembrane channels that serve as pumps to expel fluoride anions which passively enter the membrane in the form of HF, out of the intracellular enviromnent. Fluoride antiporters are found in several bacteria [104], however, *E. coli* uses an unrelated channel protein termed CrcB to achieve fluoride resistance. Overexpression of the *crcB* gene augments chromosome condensation, when expressed along with one or two other genes, leading to increased camphor resistance which ranges from 10 to 100-fold with respect to levels which are lethal to organisms under normal growth conditions [106], [107]. That previous functional focus probably led to the delay in tying this gene to fluoride resistance, which was accomplished only after the discovery of the upstream fluoride-specific riboswitch controlling its expression. The mutant *E. coli* strain in which the *crcB* gene is deleted, exhibits a 200-fold less minimum inhibitory concentration of fluoride than the wild type. The effect of other halides remains unchanged indicating the very specific nature of this channel to export fluoride [9]. The *crcB* motif is also encountered in eukaryotic organisms where thousands of homologues exist [105]. In most occurrences throughout bacterial homologues, CrcB is a small dual-topology protein of around 100 amino acids comprising four transmembrane

subunits, and the active channel is formed by a dimer with the two monomers aligning in opposite orientation through the membrane [108], [109]. On the other hand, several eukaryotic homologues exhibit a monomeric structure that consists of two homologous domains probably evolved from a gene fusion event. Later, it was shown that only one of the two fluoride-accepting pores was active in the monomeric structure indicating an evolved inactive pore in one of the homologous domains [110].

4.2.3 Intracellular availability of fluoride

As mentioned above, the *E. coli* bacterium in its wild-type contains a FEX (fluoride export) protein, termed CrcB, which highly increases its viability in fluoride concentrations as high as 200 mM. On the contrary, knockout (KO)s of *crcB* in this bacterium exhibit approximately 200-fold more sensitivity in the fluoride anion showing severe inhibition of growth in micromolar concentrations and complete absence of growth in little more than 1 mM. The fluoride exporting functionality of the protein is indirectly evidenced by the activation of the *lacZ* gene placed downstream of the *Pseudomonas syringae* $eriC^F$ fluoride riboswitch [9]. The protein exhibits the same dimeric molecular structure as encountered in many homologues and is highly efficient in exporting fluoride with a k_{cat} of at least $3x10^4\ s^{-1}$ as determined by kinetic studies [111].

The creation of an *E. coli* host based on the objectives set out in the beginning of the chapter should follow a very simple rationale: The availability of substrates in the intracellular medium. With this in mind, studying and the internal mechanisms of bacteria that already possess the ability to elaborate fluorometabolites *in vivo* could provide very useful insight. An initial study with respect to *S. cattleya* showed that the fluoride uptake - and therefore intracellular availability - is directly influencing the *in vivo* fluorometabolite production [8]. The severe

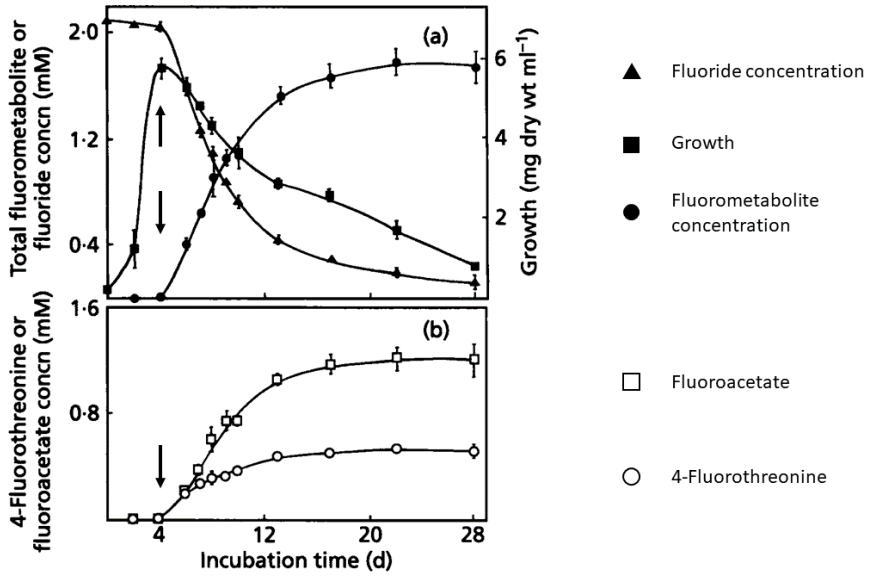


Figure 4.6: Figure adapted from [8] showing fluoride consumption and fluorometabolite biosynthesis of *S. cattleya* incubated in 2 mM fluoride.

impediment in growth with little more that 2 mM of potassium fluoride (4 days instead of 2 for full growth) suggests that this bacterium does not include the same fluoride export mechanism that other bacteria do, such as *E. coli crcB* or *P. syringae eriC*^{F}. On the contrary, an interpretation of the kinetics observed in that same study, suggests that this Gram-positive bacterium is adapting a fluoride uptake that is only expressed once the bacterium achieves maximum growth. This putative fluoride uptake system has was also found to be dependent on the pH of the growth medium. Although the fluoride uptake system has also been assumed after the discovery and characterization of the fluorinase enzyme [112], there is no evidence to date suggesting the existence of a channel or transporter that mediates fluoride uptake.

An alternative interpretation, which is also strengthened by pH dependence, is that the fluoride efflux mechanism is simply absent in *S. cattleya*, hence the severe inhibition of growth in 2 mM of fluoride mentioned in [8]. Fluoride can passively

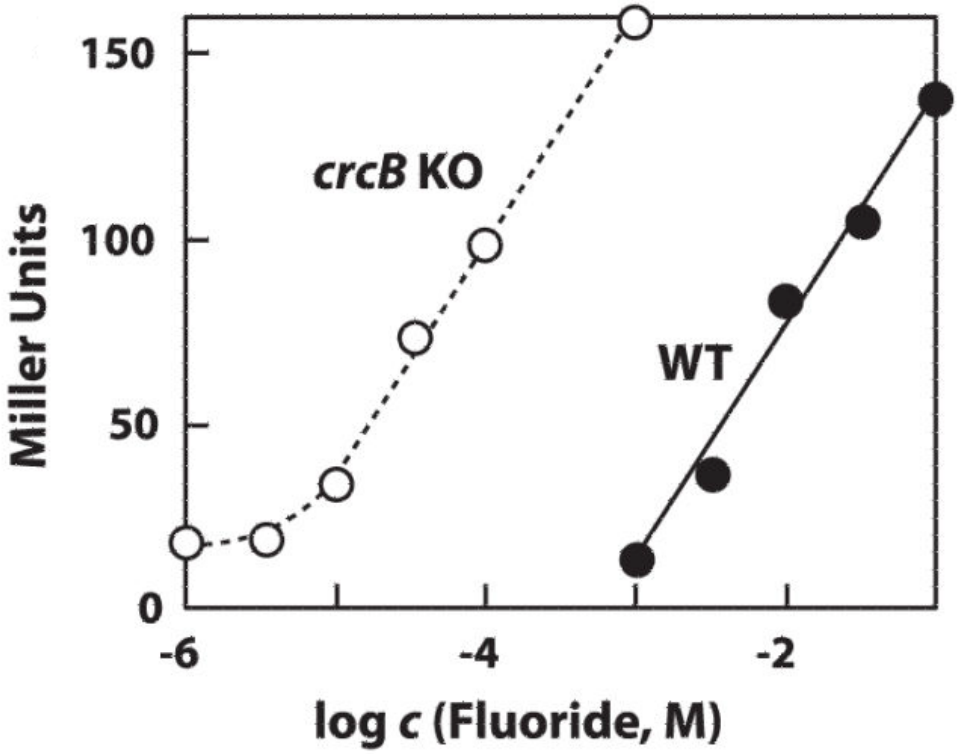


Figure 4.7: Figure from [9] showing fluoride-dependent activation of the reporter gene fused with the *crcB* riboswitch.

enter the membrane in HF form by diffusion, a mechanism which is found to be strongly dependent on pH [113]. The aforementioned study reveals that the fluoride uptake mechanism in fluc-absent *E. coli* bacteria does not involve a particularly evolved system for fluoride influx but rather simply a passive diffusion of HF molecules which are membrane permeable. Moreover, under lower pH values, the influx and accumulation of fluoride intracellularly is higher leading to a lower value for the ratio of extracellular to intracellular fluoride concentration, which could be misinterpreted as the result of an active fluoride uptake mechanism. The sudden increase in the rate of fluoride consumption in *S. cattleya* after reaching full growth observed in figure 4.6 could be attributed to the induction of the pathway leading to fluorometabolite synthesis which starts

consuming the intracellular fluoride resulting in an analogous rate of passive fluoride uptake in order to maintain equilibrium.

Combining the above research facts, the deletion of the *crcB* gene, encoding for the fluoride efflux pump was deemed necessary as a natural second step towards creating an *E. coli in vivo* fluorinating host, since it is logical that if one of the substrates of the fluorinase is being readily exported from the intracellular environment, the reaction would not be able to start. The high k_m constant for fluoride which lies in millimolar levels [41], [1], is also a factor which contributes towards the rationale of this approach, as *E. coli* cells with a functioning *crcB* gene grown in 100 mM of NaF, exhibit activation of a reporter gene bound to a fluoride riboswitch resulting in lower intensity than *E. coli* cells with a KO *crcB* genenotype grown in 1 mM of NaF (figure 4.7).

4.2.4 Creation of the *crcB* gene KO in BL21(DE3)

In this section, the creation of a *crcB* KO of BL21(DE3) is described. To the best of the author's knowledge, this is the first attempt at KO of *crcB* for this genotype and was performed as a continuation step for the strain development towards *in vivo* fluorination from the fluorinase expressing *E. coli* BL21(DE3) presented in section 4.1. The method of deletion is described in detail in section 2.4.1. This particular approach was adopted over others for a variety of reasons. First of all, it is a scarless method upon completion of the necessary steps. Once the KO genotype is confirmed no other steps are necessary, while other more popular methods often leave "scars", for example antibiotic resistance cassettes. Since most researches will not make the extra effort to remove these "scars", because it is not necessary for the illustration of a particular result tied to the respective KO, during distribution of research material there are often undesired or undocumented scars which hinder the research efforts of the next

researcher. Another reason for this approach is that it allows for multiple edits of the chromosome, without the need for removal of left-over markers. Finally, the utilization of electroporation systems or temperature dependent expression is optional, steps which are rather sensitive in other methods such as recombineering [61].

The deletion of *crcB* follows the protocol first published by Posfai et.al [114], which shows the process in *E. coli* and later adopted for other Gram-negative bacteria by de Lorenzo group [115]. The first step towards creation of the knock-out, is the construction of a particular type of plasmid, which carries homologous regions of 600-800bp upstream and downstream of the gene to be deleted. In-between the regions, instead of the full gene, two codons are included, a start and a stop codon. This is referred to as the deletion genotype and ensures an in-frame deletion to avoid polar effects. These plasmids should have a particular characteristic: they must not be able to replicate in the deletion host. Assuming that the plasmid contains a particular antibiotic resistance cassette, the clones which survive and grow in this antibiotic are the ones where the homologous regions have been recognized by the recombination machinery of the host and have integrated the plasmid in the chromosome. The initial method by Posfai utilized plasmids than carry either a temperature-sensitive origin of replication or an R6K replicon. The vector with the temperature sensitive replicon will not replicate above 37 ^{o}C, while the R6K replicon needs the *pir* gene in order to replicate. In both cases, plasmids can be maintained in cloning strains either with growth below 32 ^{o}C or in particular strains which carry the *pir* gene. The process of plasmid integration is shown in figure 2.10.

The backbone plasmid that was used for the creation of this suicide vector is the pSEVA612S [116] and was kindly provided by Victor de Lorenzo in a cc118-λ*pir* strain. The plasmid isolation was challenging because of the nature of this strain which contains DNAses (see table 4.7). The problems have risen in particular

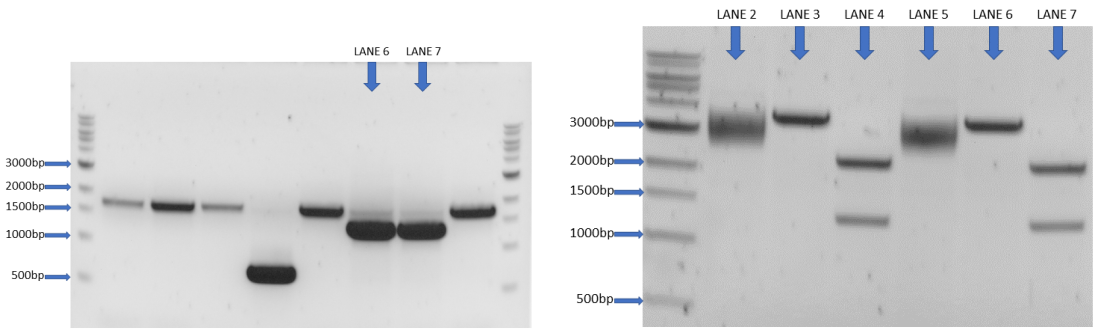
P1_DCRCB_FW	CG-GGATCC-AGCTCCCTGTAATGCCTGTG
P2_DCRCB_RV	TTTGGTCT-TTACAC-TGCAAAATCCTGCTATTTGA
P3_DCRCB_FW	ATTTTGCA-GTGTAA-AGACCAAAAAAAAACCCGCT
P4_DCRCB_RV	GC-TCTAGA-AAAGGAAACTGGCATGGCCTG
P1_C_VAL_FW	AGGGATCTATAATCCGGCTTTGC
P2_C_VAL_RV	CATGCACGTTTCGCTTACGAC
T7_VAL_FW	ACGACCGTTAGCCCAATAGC
T7_VAL_RV	TCTCTGACATCGAACTGGCTG

Table 4.3: Primers for the construction/sequencing of the pSEVA612S-*crcB* deletion plasmid and for validation of integration in the BL21(DE3) chromosome. Red colour signifies restriction sites, brown colour corresponds to sequences inside homologous regions upstream and downstream of *crcB*. The green colour shows sequences that are part of genes.

during the initial extraction which ended up either with no visible bands before or after the restriction step, probably due to DNAse activity. After a lot of repetitions, the problem was solved by subjecting the eluate from the extraction to an extra purification process, using a gel/PCR purification kit, which presumably removed the DNAses.

The plasmid containing the deletion genotype was created as follows:

Four primers were designed. One forward primer starting at the beginning of the leftmost homologous region including a *Bam*HI restriction site (P1_DCRCB_FW in 4.3). One reverse primer containing the immediate upstream region of the *crcB* gene plus the start codon (GTG) and an 11bp tail which consists of the stop codon (TAA) and 8bps into the *crcB* downstream region (P2_DCRCB_RV). One forward primer with the stop codon plus the immediate region downstream of *crcB*, including a tail with the start codon and 8bps into the upstream region of



(a) **(b)**

Figure 4.8: (a) Amplification of the pSEVA612S with the insert of *crcB* deletion locus in lanes 6 and 7. (b) Restriction of the plasmids with *Bam*HI and *Xba*I show the expected sizes of the inserts.

crcB (P3_DCRCB_FW). Finally, a reverse primer at the ending of the rightmost homologous region with an added *Xba*I restriction site(P4_DCRCB_RV).

Two separate PCR reactions were prepared with pairs of primers in the sequence described above (P1_DCRCB_FW, P2_DCRCB_RV and P3_DCRCB_FW, P4_DCRCB_RV), resulting in two amplified fragments of 584 and 730 base pairs respectively. After purification of these fragments they were added to a new "fusion" PCR reaction. The difference in this alteration of the procedure is that instead of one template, two templates with an overlapping region are added. The rationale is that during denaturation and re-annealing, the overlapping regions will anneal and extend to create the complete template, which comprises the deletion genotype. Primers P1_DCRCB_FW and P4_DCRCB_RV are also added to the reaction in order to amplify the complete template. The tails included in the second and third primer were designed in such a way that the overlapping region consists of 22 bps and exhibits the same T_m as the pair of primers added. The resulting PCR product with the deletion phenotype was purified and subjected to a restriction reaction with enzymes *Xba*I and *Bam*HI. The pSEVA612S plasmid was also subjected to restriction with the

same enzymes in a separate reaction. To avoid non-specific digestion of the fragments, the reactions were left to incubate in room temperature for 2 hours. After purification of the reaction, the fragments were put together in a ligation reaction and left to incubate overnight in room temperature. The reaction tube was then placed in 65 ^{o}C to inactivate the ligase and then used for transformation of competent *E. coli* cc118-λ*pir* cells. The pSEVA621S is a gentamicin(Gm) resistant plasmid, therefore trasformed colonies were selected in a Gm plate and Gm - resistant colonies were picked and subjected to colony PCR with primers P1_DCRCB_FW and P4_DCRCB_RV. The resulting gel is shown in 4.8a. Successfully created (and transformed) plasmids correspond to lanes 6 and 7. The bands appearing in other lanes, apart from lane 4, correspond to amplification of the chromosomal region which contains the *crcB* gene, hence appear higher due to larger size of the amplicon, whilst the lower sizes appearing in lanes 6 and 7 correspond to amplification from the plasmid fragment missing the *crcB* gene. To further validate the insertion of the deletion phenotype, plasmids were extracted and subjected to a restriction reaction with the aforementioned enzymes. The resulting gel is shown in 4.8b. Lanes 1,2,3 and 4,5,6 are duplicates. Lanes 2,5 correspond to uncut plasmid, lanes 3,6 correspond to cut plasmid with 1 restriction enzyme and lanes 4,7 correspond to restriction with both enzymes, where the 2 fragments are of the expected size. Moreover, for the definitive validation of the plasmid sequence, both positive plasmids were sequenced and the one corresponding to lane 7 was found as an error-free match.

The next step involves the integration of the constructed pSEVA612S-Δ*crcB* plasmid in the chromosome. Two separate methods of plasmid delivery are described in [116], namely electroporation and conjugation. The latter was chosen as the most efficient method of these two. Conjugation is a process in which cells bearing the F plasmid (F^+) create a pilus and connect to other bacteria which do not bear the plasmid without the plasmid (F^-) making the transfer of genetic material possible. In this study, a form of conjugation called

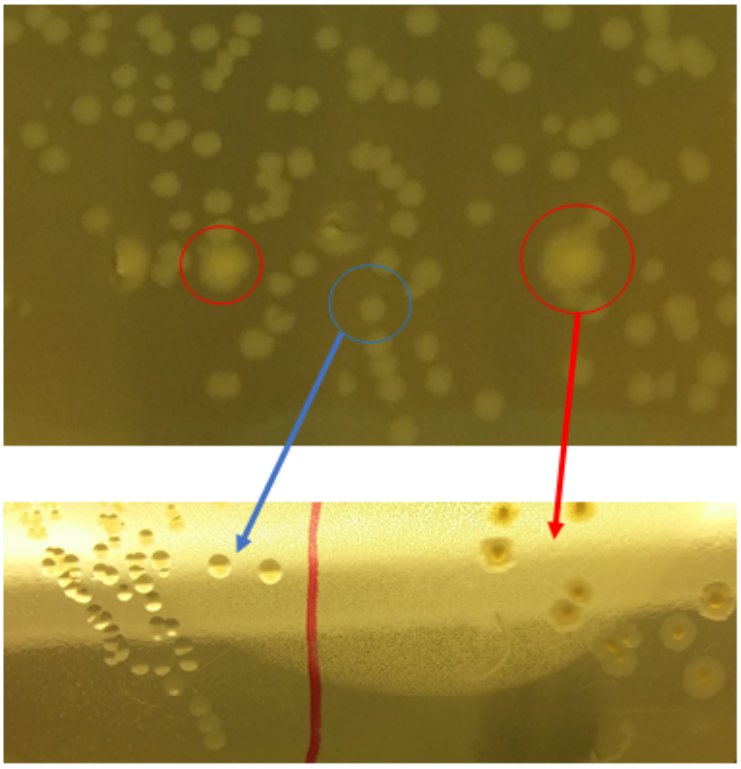


Figure 4.9: Red circles correspond to the "fried-egg" (L-form) phenotype. All bacteria from the plate showing this morphology were tested with PCR and found to be recombinants (and T7 polymerase positive), whilst all others (blue circle) were found to be T7 negative and non-recombinants.

triparental mating is used, in which apart from the donor and the recipient strain, a helper strain is also used to assist the transfer of genetic material from the donor to the recipient. In essence, the bridging between the recipient and the donor and the transfer functions are provided by the helper. The origin of transfer (*oriT*) present in pSEVA612S enables the mobilization of this plasmid from the conjugation machinery. This variant of conjugation is useful, as it is often undesirable to maintain cloning strains in a conjugative (F^+) form. The implementation in this work is an alteration to the original procedure described in [116]. In particular, the aforementioned study describes delivery of the plasmid to *Pseudomonas putida* cells which is the recipient (R) strain, with the donor

(D) and helper (H) strains being *E. coli*. This makes the final selection step rather straight-forward with cultivation of the final conjugation mixture in M9 minimal medium with citrate as a carbon source which cannot be metabolized by *E. coli*. In this implementation all strains are *E. coli* complicating the final selection step. For this process to work with BL21(DE3) recipient cells, the antibiotic resistance marker of p*flA*1-his plasmid already existent in this strain was used, along with the subtle observation that there is a visible difference in the phenotype of *E. coli* BL21(DE3) compared to cc118-λ*pir*. Initially, the conjugation mixture was prepared as described in [116] but in the final selection step, an LB agar plate both gentamicin and ampicillin was used instead of an M9 + citrate. The pSEVA612S plasmid contains an *oriT* which permits the plasmid to be transfered through the conjugative bridge, while p*flA*1-his does not contain the *oriT* sequence, therefore in theory, using an Ap + Gm plate the only colonies that should appear are BL21 (DE3) with the additional Gm resistance from the pSEVA612S that managed to move from cc118-λ*pir* to BL21 (DE3) and integrate to the chromosome of the latter. Along with the full conjugation mixture (D+H+R), the following combinations were also tested: D, H, R, D+R, D+H, H+R in Ap+Gm plates. Apart from the plates corresponding to D+H+R, colonies appeared in the plate corresponding to D+R as well, probably due to rare events of p*flA*1-his plasmid uptake by cc118-λ*pir* cells. In the D+H+R plate however, there was a visible difference in some colonies as shown in 4.9. A number of those which exhibited the different form were restreaked and undergone colony PCR. To our surprise, all colonies with the different morphology were cointegrates, while colonies that had the same morphology as in the D+R plates, were not. This result indicates that it is possible to select cointegrates by visual inspection of the selection plate. The pSEVA612S plasmid is inserted as a linear fragment with the homologous region residing either in the beginning or in the end of the fragment, based on the position of insertion to the chromosome. Depending on which region of the chromosome is recognised by the recombinases (upstream

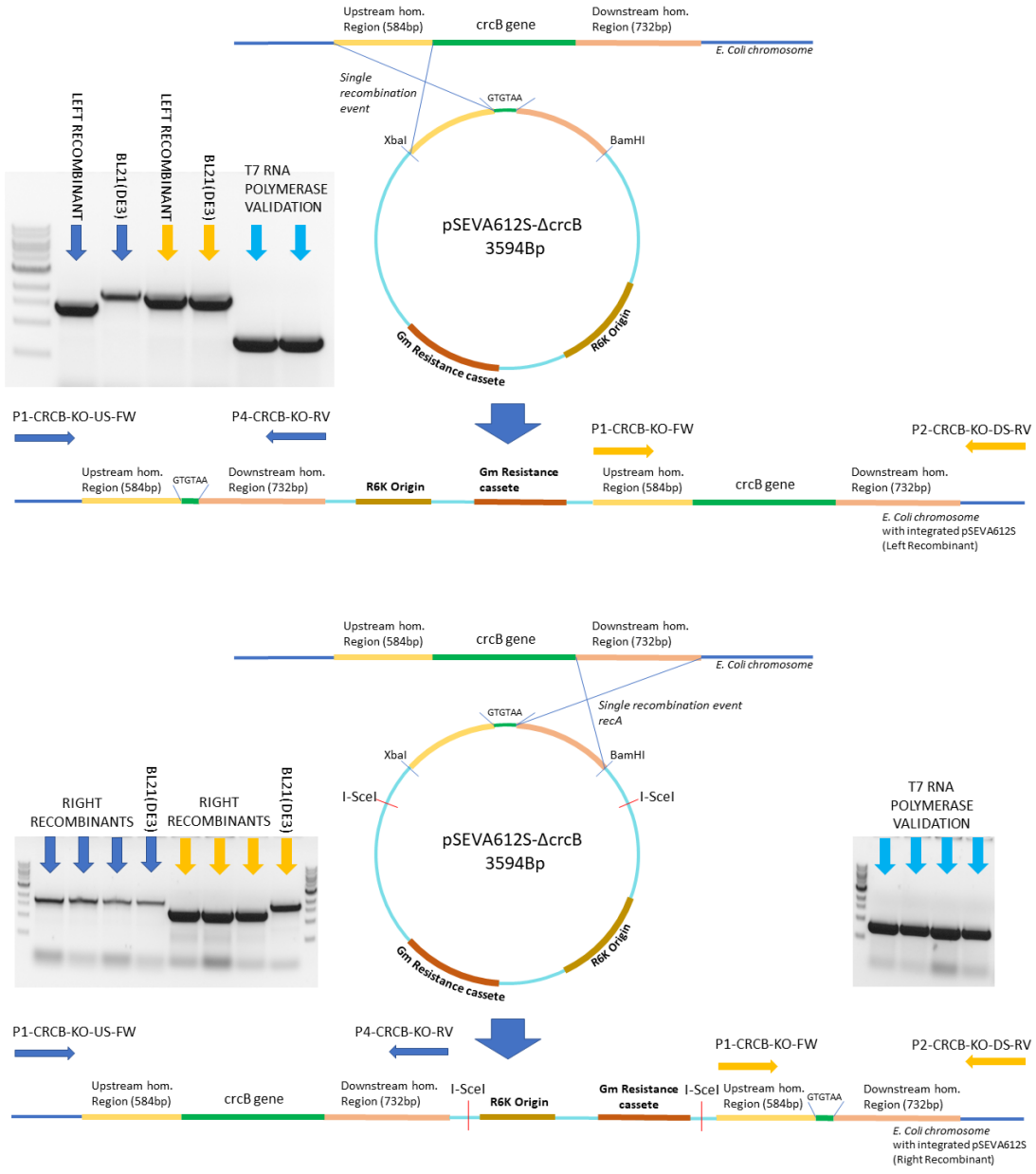


Figure 4.10: Integration of pSEVA612-Δ*crcB* plasmid upstream or downstream of the target locus. Left Recombinant (Top) and Right Recombinant (Bottom) with the respective primers and validation gels.

or downstream) the plasmid is integrated either upstream (left recombinant) or downstream of the chromosome's respective region (right recombinant). To confirm the integrate' s position, primers were designed upstream (P1_VAL) and downstream (P2_C_VAL_RV) of the homologous region. Each colony was tested by PCR with two pairs of primers (P1_VAL, P4_DCRCB_RV) and (P1_DCRCB_FW, P2_C_VAL_RV). The resulting gel and the sizes of the bands corresponding to these pairs determine the position of the integrate in the chromosome as shown in figure 4.10. The cointegrates were further tested to validate that they were BL21(DE3) and not cc118-λ*pir* cointegrates, although this would be highly unlikely in cc118 cells due to the modified *recA1* genotype modification which reduces recombination efficiency. The main difference between BL21(DE3) and cc118-λ*pir* is the T7 polymerase and this has been tested with a number of cointegrate colonies, with primers designed inside the T7 polymerase gene. As shown in figure 4.10, 3 left recombinant, 1 BL21(DE3) unmodified and 4 cc118 colonies were tested for amplification of the fragment corresponding to T7 polymerase and all of them showed the expected result. Following confirmation of the integrate's location, colonies were isolated corresponding to both right and left recombinants and chemical competent cells were created. These were then transformed with the kanamycin resistance plasmid pSEVA228S which contains the meganuclease I-*Sce*I enzyme under the control of the xylS/Pm promoter. After introduction of the plasmid, the left and right recombinants were grown as a liquid culture in LB with kanamycin and expression of the I-*Sce*I enzyme was induced with the addition of m-Toluic acid to a final concentration of 1 mM. This step involves the creation of two DSBs by the enzyme in the two I-*Sce*I sites present in the integrated fragment. These DSBs in turn induce a SOS response which is resolved either by homologous recombination leading to one of the genotypes present (wild type or Δ*crcB*), or by cell death. This step mediates an effective selection for the resolutes as well. This deletion method was first illustrated with the presence of one DSB in the chromosome [114]. However,

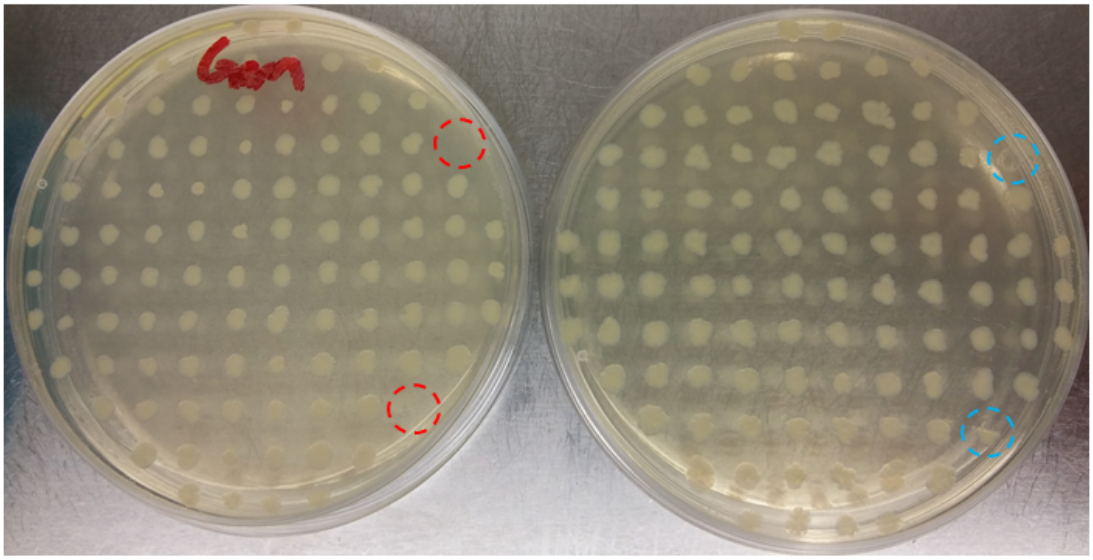


Figure 4.11: Negative selection of Gm sensitive clones. Two colonies (red circles) are missing in the Gm plate and the corresponding colonies (blue circles) in the plate w/o antibiotics were picked as resolutes.

according to [115], the addition of one more I-*Sce*I site increases DSB resolution efficiency. The surviving resolute cells will be Gm sensitive because the DSBs will have released the Gm resistance cassette fragment from the chromosome, hence Gm is not added during the induction of I-*Sce*I cleavage. Following induction of cleavage, appropriate dilutions of cells were prepared in agar plates without antibiotic. After overnight growth, 100 colonies (50 from left and 50 from right recombinants) were re-streaked in two plates, one without antibiotic and one with Gm. Each of the 100 colonies was isolated in both plates corresponding to the same numbered position. This is the negative selection step which would indicate which colonies are Gm resistant and which not, with the latter corresponding to resolutes. A dilution from the previously grown liquid culture was passed into a new tube with LB, kanamycin and 1 mM of m-Toluic acid for another round of overnight growth and colony picking. After 6 passes, two distinct Gm-sensitive colonies appeared as shown in figure 4.11. Both colonies were subjected to PCR with the pair of primers P1_C_VAL_FW and P2_C_VAL_RV and were found

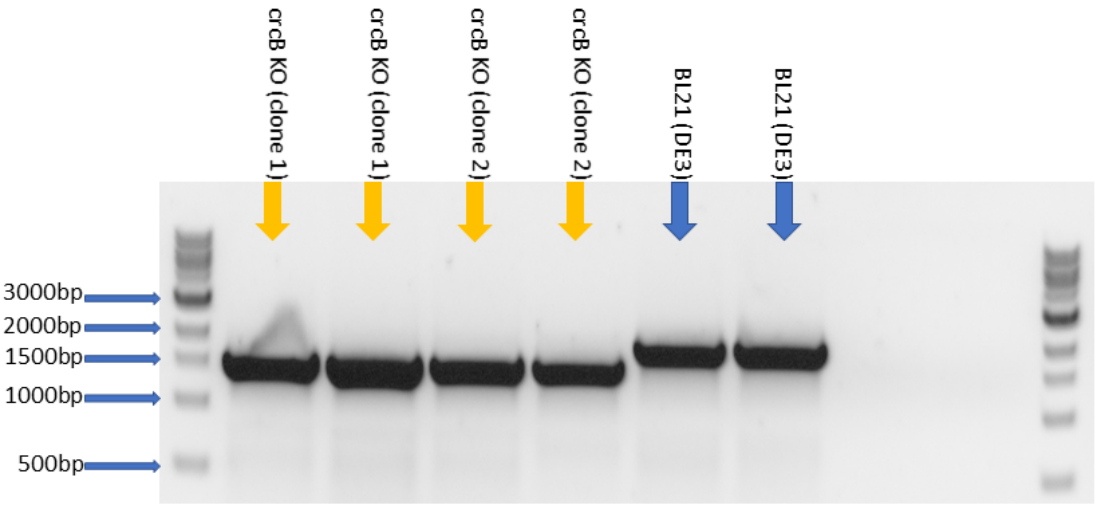


Figure 4.12: PCR validation of the KO genotype in the resolutes picked as indicated in figure 4.11.

to contain the *crcB* KO genotype as shown in figure 4.12. In order to validate the *crcB* KO phenotype, these colonies were re-streaked and placed in liquid LB with 3 mM of KF, in which no growth was visible after 2 days. Furthermore, to cure plasmids that would still potentially reside in this strain, such as p*flA*1-his or pSEVA228S, the strain was put in a liquid culture and left to grow overnight, diluted and grown again in several passes. Then colonies were put in plates with no antibiotic, Kn and Ap separately and those found to be sensitive in both antibiotics were picked.

The number of passes that were needed until the two *crcB* KO positive clones were obtained is much closer to the spontaneous recombination rate of $1.7x10^{-3}$ [114]. A much more robust resolution of the cointegrates was expected, indicating that the DSB from cleavage of the I-*Sce*I sites was not properly induced. A number of possible reasons could lead to this effect. Incomplete induction of expression of the I-*Sce*I meganuclease, plasmid instability of pSEVA228S or absence of the correct I-*Sce*I sites from the integrate are some of them. The outcome of the

method however did not change, despite this set-back, fact which constitutes one more argument for its choice.

4.3 Purine nucleoside phosphorylase and adenosine degradation

PNP is a term which describes a variety of enzymes which, as the self-explanatory name suggests, accept a purine nucleoside, namely adenosine, deoxyadenosine, inosine, deoxyinosine, guanosine, deoxyguanosine or chemically similar compounds, and a phosphate as substrates yielding α-D-ribose 1-phosphate and the respective base. In several organisms, it is encountered as part of the purine salvage pathway which mediates the production of purine monophosphates as an alternative to normal purine synthesis, where the latter is compromised or missing in the first place. In higher mammals (rats, humans), PNP is expressed in many different tissues but is mostly found in the liver and is produced by hepatocytes. Mutations in the PNP gene lead to an disorder known as PNP deficiency which is autosomal recessive. Individuals with this disorder show symptoms of CNS impairment and immunodeficiency. Human PNPs significantly differ from bacterial ones structurally and although they accept the same substrates in their key function in purine metabolism and mediate the same reactions, bacterial PNPs exhibit a broader substrate specificity, fact which led recent endeavours towards the development of a suicide gene therapy for cancer treatment. The main theme behind this strategy is the activation of prodrugs by cleavage from *E. coli* PNP but not from its human counterpart. By expressing this PNP in cancer cells, toxicity of the activated prodrug which is otherwise inert in normal tissue, kills off the cancer cells. [117].

4.3.1 The purine nucleoside phosphorylase as part of the 4-fluorothreonine pathway

The second step in the fluorometabolite biosynthesis pathway of *S. cattleya* is mediated by an enzyme which accepts 5'-FDA, the product from the direct fluorination reaction, and a phosphate, producing Adenine and 5-FDRP, therefore it is a classic PNP. Evidence for this step were found by incubating cell free extracts of the aforementioned bacterium in the presence of 5-FDRP which yielded the final product of the pathway, suggesting that the 5'-FDA is processed by a PNP [42]. Sequence similarities and prediction by homology, suggested that the ORF directly upstream of *flA*1 corresponded to this putative PNP and definitive proof was provided by assays with the product of this ORF. In particular, the partially purified product from cell-free extract performed phosphorolysis of 5'-FDA yielding 5-FDRP [47]. The gene corresponding to this enzyme was therefore assigned the name *flB*. In four other occurrences of the fluorometabolite pathway (*flB*, *flB*1-*flB*4), the corresponding gene is either directly adjacent or very close to the related fluorinase gene and with sequence similarities that range from 59%-74%. In most of these cases, where the pathway is active (i.e. producing fluoroacetate) [53], [55], position and similarities of the enzyme suggest that it has been evolved specifically to process the product of the nearby fluorinase gene. Evolution drift, responsible for the inactivation of the pathway in *N. Brasiliensis* or *Actinoplanes* sp. N902-109, has not displaced this gene much further away as is the case with other genes of this pathway. It is worth noting that the enzyme is also a catalyst for adenosine although with a much lower efficiency than its fluorinated analogue.

4.3.2 The *E. coli* purine nucleoside phosphorylase(s)

There are two different PNPs present in the wild type *E. coli*, and the respective genes are termed *deoD* and *xapA*. The product of *deoD* is a homohexamer with a broad substrate specificity. Although its function as part of the purine salvage pathway processing of inosine to yield hypoxanthine leading to degraded purines, there is also high reactivity with adenosine, guanosine and some interesting fluorinated substrates such as 2-fluoro-adenosine (F-Ado) and 2-fluoro-2-deoxyadenosine (F-dAdo) [118] but not xanthosine which is processed by the product of *xapA*. On the other hand, *xapA* cannot process adenosine or deoxyadenosine but can process inosine, guanosine and their deoxy-counterparts. Interestingly, catalysis of the phosphorolysis of the fluorinated purine nucleosides F-Ado and F-dAdo by the DeoD-type PNP, exhibits higher affinities and efficiencies attributed to stronger van der Waals bonds with the respective active site residues [119], [120]. The product of *deoD* (termed *pup* in earlier studies) is vital for the metabolism of purine nucleosides such as adenosine and deoxyadenosine and mutants for this gene do not grow in the presence of each one of these metabolites as sole carbon sources [121].

4.3.3 Rationale and creation of the *deoD* KO in *E. coli* BL21(DE3)

The broad spectrum of substrates accepted by both of the PNP enzymes present in wild type *E. coli* but especially the compatibility of the DeoD-type PNP with fluorinated analogs of adenosine and deoxyadenosine indicates the possibility that other fluorinated analogs such as 5'-FDA might be degraded as well by this particular phosphorylase. On the other hand, XapA type expression is only induced in the presence of xanthosine [120] and xanthosine is not produced up

P1_DDEOD_FW	CG-GGATCC-TGTTTACCGACAGCCTCAACC
P2_DDEOD_RV	AGGATAAAACA-ATGTAA- TTGTGTTTCGCTGAAAGGCG
P3_DDEOD_FW	GCGAAACACAAT-TACATT- GTTTTATCCTTATCAACATGTTTTCACCC
P4_DDEOD_RV	CG-GAATTC-TGACTCCGTGTTCCAGATTGC
P1_D_VAL_FW	TGGATGAACGCTTTAGCTGGG
P2_D_VAL_RV	CTCTTCCGGTACGGTCATTCTG
M14_612_FW	CAGGAAACAGCTATGACCATG
M14_612_RV	TGTAAAACGACGGCCAGT

Table 4.4: Primers for the construction/sequencing of the pSEVA612S-*deoD* deletion plasmid and for validation of integration in the BL21(DE3) chromosome. Red colour signifies restriction sites, brown colour corresponds to sequences inside homologous regions upstream and downstream of *crcB*. The green colour shows sequences that are part of genes. Blue colour indicates sequences that reside in the pSEVA612S plasmid (used for sequencing).

to less than micromolar scale by *E. coli* growing in a standard glucose-based medium. Adenosine (or deoxyadenosine), an inducer of *deoD* [122], is however detected during growth of *E. coli* [123]. Moreover, a recent study for the directed evolution of the fluorinase reported the deletion of the *E. coli* PNP enzyme to avoid degradation of the product 5'-FDA during high-throughput analysis of cell-free extracts [52]. Although no ample justification or data supporting this position were presented in the latter, it was determined that deletion of *deoD* for the avoidance of 5'-FDA degradation would be a prudent course of action nonetheless.

The same methodology was utilized for the deletion of the *deoD* gene as described in section 4.2.4, with a modification in the step of the suicide plasmid delivery. For completeness of the strain modifications included in this work, both *E.*

coli BL21(DE3) and its Δ*crcB* counterpart were subjected to this procedure. Primers were designed as before (see table 4.4) with appropriate restriction sites (*Eco*RI and *Bam*HI) and the fusion template including the Δ*deoD* genotype was amplified. The resulting linear construct and plasmid pSEVA612S were subjected to separate restriction reactions. After restriction and purification from the reactants, plasmid pSEVA612S and the linear fragment of the *deoD* KO genotype were confirmed with electrophoresis. Ligation of these fragments followed, resulting in the desired pSEVA612S-Δ*deoD* plasmid, which was delivered to cc118-λ*pir* cells by chemical transformation. Colony PCR was performed in the transformed bacteria and gel visualization indicated correct sizes in most colonies. Plasmid extraction was performed from 4 positive colonies and a restriction with the pair *Bam*HI, *Eco*RI followed to further validate the correct insert sizes. Sequencing was subsequently performed in two of these plasmids yielding complete correspondence with the theoretical sequence.

For plasmid delivery in the strain and integration to the chromosome, the previously used method of conjugation could not be repeated for strain MK009 (table 4.8) because of the acquired phenotype F^+ from the previous deletion, which forbids conjugative transfer of plasmids to these strains. Therefore, chemical transformation was used. The caveat of this approach is that chemical transformation efficiency is up to 1000-fold less than electro-transformation. For this reason, the reaction masses of both cells and DNA were scaled up 20-fold, while incubation time was increased by a factor of 2, with respect to chemical transformation for normal plasmid delivery. A small number of left and right recombinant colonies were obtained for both the unmodified BL21(DE3) and the MK009 strain. Genotypes were confirmed with gel electrophoresis as with the previous deletion with appropriate primers designed upsteam and downstream of the homologous regions, paired with the respective forward or reverse primer. As a next step, one recombinant from each strain was transformed with pSEVA228S to mediate the two DSBs in the I-*Sce*I sites. Based on the experience of the

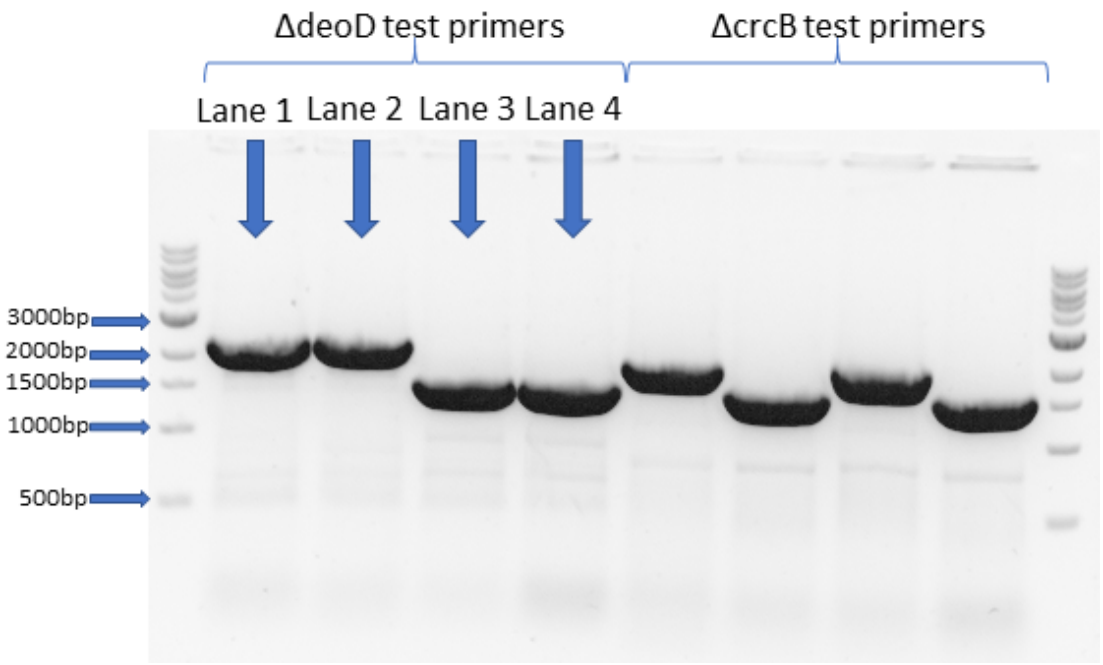


Figure 4.13: Strains BL21(DE3) (Lane 1), MK009 (Lane 2), MK021 (Lane 3) and MK017 (Lane 4) amplified with primer pairs P1-DEOD-KO-US-FW, P2-DEOD-KO-DS-RV (Δ*deoD* test primers) and P1_C_VAL_FW, P2_C_VAL_RV (Δ*crcB* test primers). All combinations of KOs are confirmed.

previous deletion which required 5 cycles of growth until 2 Gm-sensitive colonies appeared, in this instance, after induction of the xylS/Pm promoter with m-Toluic acid for I-*Sce*I meganuclease expression, the strain was subjected to 5 cycles of overnight growth. Restriction of colonies to Gm and antibiotic-free agar plates showed that all colonies were Gm sensitive. This was a first indication that the meganuclease worked as expected contrary to the previous deletion round. Gel electrophoresis did not reveal any KO genotypes after two repetitions of the process. The incubation was then run for only one overnight cycle. Again almost all colonies were Gm-sensitive but this time a number of colonies with KO-genotypes were obtained however, much less than 50 percent. This is an indication that the Δ*deoD* phenotype grows at a disadvantage to unmodified strains and is also the reason why incubation for more cycles results in dilution

of the KO-genotype colonies. The two strains bearing the deletion Δ*deoD* were then grown in the absence of antibiotic to cure the pSEVA228S plasmid as before. Finally, they were tested with colony PCR along with the BL21(DE3) strain as shown in figure 4.13 where strains Δ*deoD* and the double deletion Δ*crcB*Δ*deoD* are illustrated. To the best of our knowledge, the latter is the first implementation of this double KO strain.

4.4 SAM transporter expression and the full-mod strain

The final step of the creation of the modified host for *in vivo* fluorination will be described here, namely SAM transporter expression. The work that has been done will be described and an explanation of how the expression of this membrane protein will enable the increase of intracellular SAM by its addition to the medium in higher molarities, will be included.

The final modification towards the creation of the *in vivo* enzymatic production of fluorine-containing biomolecules by *E. coli*, is the establishment of the availability of the precursor biomolecule for this reaction at high enough intracellular concentrations, namely SAM - also referred to as *S*-adenosyl methionine or AdoMet. SAM is produced by a type of enzyme known as methionine adenolyltranferase (MAT) from ATP, methionine and water in an Mg^{2+} or K^{+} dependent activity, molecules which are readily available in the intracellular space. There enzymes can be found in almost every instance of life with the exception of some parasites that uptake SAM from their hosts [124]. Moreover, the high sequence homology between bacterial and eukaryotic MAT enzymes is an indication of the general nature of the proccess in all life domains [125]. SAM is a key ingredient in a

variety of biochemical processes in living cells. It is commonly utilized in methylation of DNA, RNA, proteins, phospholipids and neurotransmitters acting as the main methyl donor in enzymes known as SAM-dependent methyltransferases. SAM is also involved in other pathways such as polyamine synthesis, yielding 5'-methylthioadenosine (MTA) and in radical-based catalysis such as biotin production.

4.4.1 The importance of SAM in *E. coli*

SAM production in *E. coli* follows the same trend described above by a MAT-type enzyme produced by the *metK* gene. Previous studies (see [126]) stated the existence of another copy of this gene in the chromosome of *E. coli* K12 adjacent to *metK* but further analysis did not corroborate this finding, indicating an isolated duplication event to the studied strains. The importance of SAM in *E. coli* is paramount, as illustrated by the fact that complete deficiency of the *metK* gene results in non-viable *E. coli* cells. Furthermore, addition of SAM in the extracellular growth medium does not ameliorate this effect, an observation that proves the lack of membrane permeability in *E. coli* for SAM molecules contrary to yeast [127]. A partially disrupted *metK* variant, termed $metK_{84}$, with an "A to G" replacement in the promoter region of the transcribed sequence, has been used in several studies for the determination of the essential SAM-dependent metabolic step whose suppression hinders growth. Regulation of the SAM pool from low to negligible levels was achieved by supplying the inducer leucine for this mutated promoter to the growth medium in different amounts. The resulting cells have shown growth defects in conditions of limiting SAM. Specifically, cells exhibited a peculiar phenotype with increased filament lengths indicating that the cell wall division process was disrupted. Restoration of normal growth was made possible by an external supplementation of the *metK* gene [128]. Later it has been shown that SAM plays a pivotal role in the assembly of a protein complex termed "septal

ring" which performs division of the wall of elongating *E. coli* cells [129]. The secondary effect of increase in the mutation rate, under limiting SAM conditions, had been proposed and initially disproved [130], however excessive limitation in SAM availability, leads to somewhat increased mutation rate by the activity of deoxycytosine methylase [131].

4.4.2 SAM as a substrate for halogenases and duf-62 proteins

Apart from the essential pathways in which SAM participates, there are enzymes directly utilizing SAM for the production of secondary metabolites. The fluorinase *flA*1 utilized in this study is a representative example. Other enzymes include chlorinase SalL from *S. tropica* which is part of the pathway leading to salinosporamide A, a bioactive compound demonstrating cytotoxic activity and used as an anticancer agent [132]. The fluorinases that have been characterized and SalL share around 35% of sequence identity and their compatibility and mechanistic relatedness is also demonstrated by the genetic substitution of the chlorinase in the *S. tropica* with the fluorinase from *S. cattleya*, resulting in the formation of the fluorinated analog of salinosporamid A, namely fluorosalinosporamid [5]. It appears that SAM-dependent halogenases are rare occurrences of the general family annotated as domains of unknown function-62 (duf-62) consisting of more than 200 proteins most of which are uncharacterized. Apart from the halogenating enzymes of this family, a number of duf-62 proteins have been analyzed and their function is the hydrolysis of SAM yielding adenosine and L-methionine utilizing an S_2N nucleophilic attack mechanism in which water acts as the nucleophile [133]. Phylogenetic studies also suggest the common evolutionary ancestry of all duf-62 proteins despite the divergence of the halogenases towards a more specific

role. The close similarity of tertiary and quaternary structure of the characterized duf-62 enzymes strengthens this hypothesis even further [134].

4.4.3 SAM transporters

Although the main source of SAM in all domains of life is synthesis from the activity of MAT enzymes, in several eukaryotic and prokaryotic paradigms, SAM permeability through the membrane has been illustrated. For example, *S. cerevisiae* has the capacity of SAM transport from the extracellular medium by specialized membrane proteins or general diffusion based systems [135]. This molecule however is not membrane-permeable in certain bacterial species, therefore, specific membrane proteins have been evolved that confer this functionality as is the case with other important biomolecules. Cellular organelles such as mitochondria, which have a distal evolutionary relation to bacteria also lack the ability to transport SAM by diffusion but rely on their host-cells for the essential supply of SAM facilitating SAM transporter functionality. Two mitochondrial examples include the *S. cerevisiae* Sam5p, part of the mitochondrial carrier family which counts 35 members in yeast genome, hence its sub-cellular location [136] and the *Homo Sapiens* SAMC which was discovered from database homology search by its sequence similarity to the yeast orthologue, although the latter does not employ an substrate-exchange mechanism as SAMC does [137]. Evidence also suggests that other hetero-metabolite dependent organelles facilitate SAM transport systems such as chloroplasts, for example the *Arabidopsis Thaliana* SAMT1, which is a counter-transport system exchanging SAM for SAH [138].

Any bacterial species that happen to develop such a metabolic route would have to exhibit a symbiotic (such as mitochondria/chloroplasts) or parasitic relation, with their host cells serving as SAM supplying pools. It therefore constitutes no surprise that species which have evolved this elaborate system are mainly

pathogens that reside in eukaryotic hosts and they did so by being exposed to an external supply of SAM. Indeed, two examples of bacterial SAM transporters that have been characterized include the *R. prowazekii* RP076 [124] and the *Chlamydia trachomatis* L2 CTL843 [139]. Both of these bacteria are pathogenic parasites and reside intracellularly in eukaryotic cells. The *R. prowazekii* has been described as the most closely related organism to mitochondria, which are thought to have been integrated to the cell cycle forming a symbiotic relation [140]. In addition to being unrelated in terms of homology, the two transporters' main functional difference, apparently common among instances of SAM transporting enzymes, is the ability of the chlamydial CTL843 to exchange SAM for SAH, whilst rickettsial RP076 performs one-way transport. The rickettsial transporter was first identified in an *R. prowazekii* strain whose genome sequence revealed a nonsense stop codon mutation in the single *metK* gene, suggesting the inability of this strain to synthesize SAM on its own. Since SAM is an essential metabolite in all life domains this finding suggested the presence of a transporter which can mediate SAM uptake from the host cell's cytoplasm. On the other hand, genomic analysis of several *Chlamydia* species showed the absence of both SAM synthesis and SAH detoxification genes. These are both examples of reductive evolution commonly seen in parasitic organisms [141]. The complementation of SAM synthesis deficiency with SAM transport was investigated in *E. coli* by the expression of both of the above transporting enzymes. The resulting $\Delta metK$ strains expressing these systems became SAM auxotrophs showing the compatibility of these transport systems with *E. coli* [142], [139]. Moreover, the dual SAM uptake/SAH exporting ability of the chlamydial transporter was illustrated by the deletion of the MTA/SAH nucleosidase, the enzyme responsible for maintaining low SAH levels. This deletion confers growth hinderance in *E. coli*, and is partially restored by the expression of the chlamydial transporter. The long filament phenotype under SAM starvation conditions described above, was

also found to be gradually restored in *E. coli* Δ*metK* strains expressing the SAM transporter with increasing SAM concentration in the growth medium [143].

4.4.4 *R. prowazekii* SAM transporter expression in *E. coli*

The decision of using the SAM transporter from *R. prowazekii* was based on qualitative rather than quantitative facts. Although the latter were also favourable for the rickettsial SAM transporter, it was difficult to make a direct comparison of kinetic properties as the plasmids, promoters, sequences, codon optimizations and growth media differ in each study [135], [139]. On a qualitative basis, the rickettsial transporter is an energy-dependent one-way transporter from the extracellular to the intracellular environment, while the chlamydial transporter is mainly a SAM/SAH antiporter, showing minute energy-dependent one-way activity. This is also illustrated by the fact that the rickettsial transporter, when expressed in *E. coli*, maintains a linear rate of SAM uptake for 20 minutes while the chlamydial transporter shows a decline in rate of uptake after the first minute. This is consistent with the SAM/SAH antiporter activity since SAH remains at low concentrations in *E. coli* by the activity of SAH/MTA nucleosidase. Additionally, optimal *E. coli* growth is achieved with an addition of 17.5 μM of SAM in the LB medium, while growth of *E. coli* Δ*metK* with the chlamydial SAM transporter is suboptimal even in the presence of 1 mM SAM. Last but not least, the rickettsial transporter displays a uniport type of activity (extracellular to intracellular only), which results in a concentration ratio (inside to outside) of higher than 10, contrary to the chlamydial transporter's antiporter functionality which at best can achieve a ratio of 1. The SAM transporter gene including a strong RBS site upstream, was ordered and synthesized as a gBlock^{TM} from IDT technologies using codon optimization resulting in the same amino acid sequence as the rickettsial transporter. The vector used for accommodating the gene is pSEVA631

P1_SAMT_FW	CG-GGATCC-CGCGGCCGCATCTAGAGTTGACGGCTAGCTCAGTCCTAGGTACAGTGTACTAGTGAAAGAGGA-GAAATACTAG-ATG
P2_SAMT_RV	TT-CTGCAG-TTATTGAGACTCAT
SAMT_SEQ_IN_FW	CGTTGGGTAGTAACGGTTGTAGG
SAMT_SEQ_IN_RV	TGGCAGCCAGGATCAAATACAG
R24_FW	AGCGGATAACAATTTCACACAGGA
F24_RV	CGCCAGGGTTTTCCCAGTCACGAC

Table 4.5: Primers for the construction and validation of the SAM transporter expression plasmid with backbone pSEVA631. The red colour signifies restriction sites, the blue colour corresponds to sequences inside pSEVA631. The brown colour shows sequences that are part of the codon optimized SAM transporter gene.

[84] featuring a pBBR1 medium copy number origin of replication and a gentamicin resistance cassette. The plasmid was extracted from an *E. coli* DH5a strain and subjected to restriction with *Bam*HI and *Pst*I. The synthesized gene was amplified with primers P1_SAMT_FW, which contains the priming sequence and a tail with a *Bam*HI site and the J23100 promoter, and P2_SAMT_RV with a tailing *Pst*I. The resulting PCR product was also restricted with *Bam*HI and *Pst*I and then put together with the restricted pSEVA631 to a ligation reaction yielding the construct pSEVA631-SAMT. Apart from primers R24_FW and F24_RV residing upstream and downstream of the fused gene, two more primers were also constructed, namely SAMT_INSEQ_FW and SAMT_INSEQ_RV as shown in table 4.5. These both reside inside the SAM transporter recombinant sequence and they are useful not only for additional coverage during sequencing, but also for validation of the presence and correct alignment of the gene inside the plasmid by PCR. In parallel, as an alternative and in order to validate the functionality of the synthesized SAM transporter phenotypically, the strain with

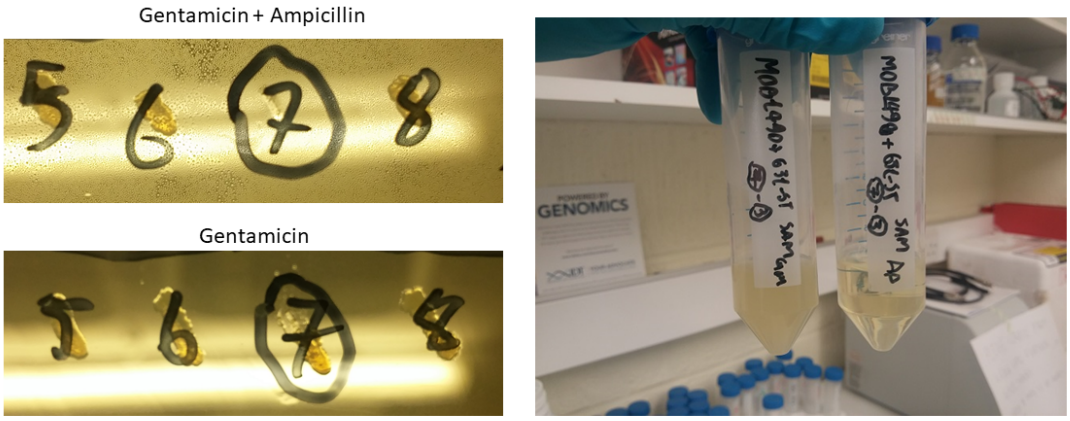


Figure 4.14: Plating in Ampicillin + Gentamicin and only Gentamicin plates. Growth of clone 7 in the latter and not in the former, shows that competitive replacement of plasmid pMW1402 with pSAMT took place successfully in strain MOB1490 (now MOB100S) validating the functioning phenotype of pSAMT.

a deleted *metK* gene facilitating the SAM transporter within the ampicillin resistant plasmid pMW1402, was kindly provided by Prof. Wood. The strain' s given designation is MOB1490 [142]. Using a negative selection replacement strategy, the plasmid pSEVA631-SAMT featuring our own version of the SAM transporter replaced the ampicillin resistant construct initially providing the SAM transport functionality. The replacement was achieved first by transforming MOB1490 to include pSEVA631-SAMT making the strain both Ap + Gm resistant. Then after several growth cycles in LB liquid (Gm), colonies were grown and isolated in LB agar plates (Gm). Several of the colonies were picked and restreaked in both Ap + Gm and just Gm plates (the same colony in the same position in both plates). The idea is that since only Gm is included during growth cycles the pMW1402 Ap resistant plasmid featuring the working SAM transporter would be diluted and lost only if the synthesized SAM transporter residing in pSEVA631-SAMT is functional. Therefore, any colonies that are Ap sensitive are the ones that have

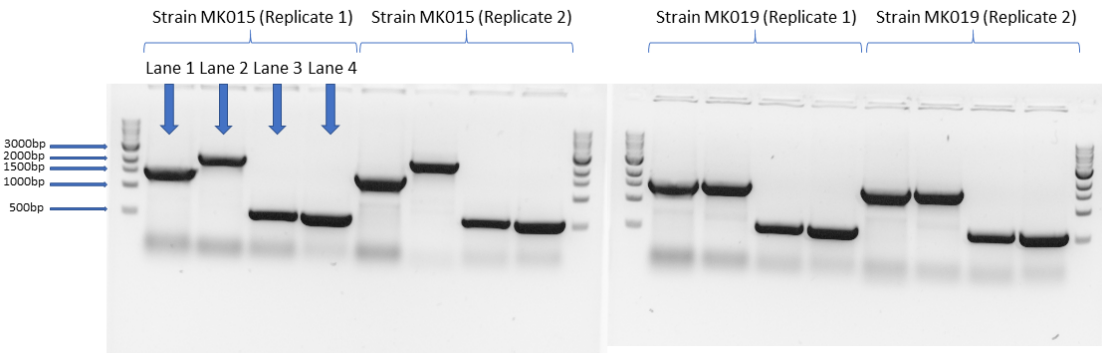


Figure 4.15: PCR products confirming all modifications. Lane 1 corresponds to primer pair (P1_C_VAL_FW, P2_C_VAL_RV). Lane 2: primer pair (P1-DEOD-KO-US-FW, P2-DEOD-KO-DS-RV). Lane 3: primer pair (pet151VAL_FW, FLA1_INSEQ_RV). Lane 4: primer pair (R24_FW, SAMT_SEQ_IN_RV).

lost the pMW1402 plasmid and these would grow on the Gm plate but not in the Gm + Ap plate. Indeed a "winner" appeared and validated by inoculation of an Ap and a Gm liquid LB tube as shown in figure 4.14. This last modification concludes the rationally chosen strategy for creating an *in vivo* fluorometabolite producing *E. coli* which, to the best of our knowledge and at the time of this development was novel. Strains that have been developed as described in previous sections were transformed with the pSEVA631-SAMT plasmid. The resulting strains MK015 (BL21(DE3) Δ*crcB* + p*flA*1 + pSAMT) and MK019 (BL21(DE3) Δ*crcB*Δ*Deod* + p*flA*1 + pSAMT) are of particular interest and a confirmatory PCR was performed in two different colonies with the corresponding gel illustrating all modifications shown in figure 4.15. Other strains with a BL21(DE3) base genotype were also constructed for in order to serve as controls and for measuring each modification's contribution to the result. All strains utilized or constructed for this work are shown in tables 4.7 and 4.8.

Plasmid	Backbone	Features	Source or Reference
pRK600	RK2013 [144]	*oriV(ColE1), RK2(mob+ tra+)* Cm^R	[116]
pJ61002	-	*oriV(colE1)* Ap^R	[72]
pSEVA612S	pEMG [115]	*oriV(R6K)*, *lacZa* fragment with two flanking I-*Sce*I recognition sites Gm^R	[116]
pSEVA631	pBBR1MCS	*oriV(pBBR1)* Gm^R	[84]
pSEVA228S	pSW-I [115]	I-*Sce*I expressing plasmid, *oriV(RK2)*, xylS/Pm → I-*Sce*I Kn^R	[116]
pSEVA612S-Δ*crcB*	pSEVA612S	Contains the *crcB* deletion homology fragment Gm^R	This work
pSEVA612S-Δ*deoD*	pSEVA612S	Contains the *deoD* deletion homology fragment Gm^R	This work
p*flA*1-his	pET151/D-TOPO	Contains the *flA*1 gene with 6xHis tag controlled by a lacl promoter Ap^R	GeneArt™
pET151(-)	pET151/D-TOPO	Same as p*flA*1-his but with the *flA*1 gene removed Ap^R	This work
p*flA*1	pET151/D-TOPO	Same as p*flA*1-his without 6xHis tag Ap^R	This work
p100-*flA*1	pJ61002	*J23100* promoter, Strong RBS, B0015 Term. carrying *flA*1 gene Ap^R	This work
p114-*flA*1	pJ61002	*J23114* promoter, Strong RBS, B0015 Term. carrying *flA*1 gene Ap^R	This work
p100-SAMT	pSEVA631	*J23100* promoter, Strong RBS, B0015 Term. carrying SAM transporter Gm^R	This work

Table 4.6: List of plasmids used in this study.

Strain	**Base Type (*E. coli*)**	**Modifications**	**Source or Reference**
MG1655	K-12 strain	F– λ– ilvG– rfb-50 rph-1	GeneArt™
DH5α	MG1655	F- λ- *endA1 glnX44(AS) thiE1 recA1 relA1 spoT1 gyrA96(NalR) rfbC1 deoR nupG Φ80(lacZΔM15) Δ(argF-lac)U169*	[116]
cc118	K-12 strain	*araD139 Δ(ara, leu)7697 ΔlacX74 phoAΔ20 galE galK thi-1 rpsE rpoB argE(Am) recAl* pSEVA228S (Kn^R)	[116]
cc118-λ*pir*	cc118	λ*pir* pSEVA612S (Gm^R)	[116]
HB101	K-12 (mostly)	F- λ- *hsdS20(rB- mB-) recA13 leuB6(Am) araC14 Δ(gpt-proA)62 lacY1 galK2(Oc) xyl-5 mtl-1 thiE1 rpsL20 glnX44(AS)* pRK600 (Sm^R Cm^R)	[116]
DH10B	MG1655	F– λ– (Str^R) *endA1 deoR$^+$ recA1 galE15 galK16 nupG rpsL Δ(lac)X74 φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC)*	Invitrogen
BL21(DE3)	B strain	F– *ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]$_{K-12}$(λ^S)*	[145]
BW25113	K-12 strain	F– λ- *lacI+rrnB$_{T14}$ ΔlacZ$_{WJ16}$ hsdR514 ΔaraBAD$_{AH33}$ ΔrhaBAD$_{LD78}$ rph-1 Δ(araB–D)567 Δ(rhaD–B)568 ΔlacZ4787(::rrnB-3) hsdR514 rph-1*	[146]
MOB1490	BW25113	Δ*metK* pMW1402 pMW1484 Ap^R Kn^R Rif^R	[142]
MOB100S	BW25113	Δ*metK* pSAMT pMW1484 Gm^R Kn^R Rif^R	This work

Table 4.7: List of strains used in this study.

Strain	Base Type (*E. coli*)	Modifications	Source or Reference
MK001	MG1655	p*flA*1-his Ap^R	GeneArt™
MK002	DH10B	pET151(-) Ap^R	This work
MK003	BL21(DE3)	p*flA*1-his Ap^R	This work
MK004	BL21(DE3)	p*flA*1 Ap^R	This work
MK005	BL21(DE3)	p100-*flA*1 Ap^R	This work
MK006	BL21(DE3)	p114-*flA*1 Ap^R	This work
MK007	BL21(DE3)	pET151(-) Ap^R	This work
MK008	BL21(DE3)	pSAMT p*flA*1 Gm^R Ap^R	This work
MK009	BL21(DE3)	Δ*crcB*	This work
MK010	BL21(DE3)	Δ*crcB* p*flA*1 Ap^R	This work
MK011	BL21(DE3)	Δ*crcB* p100-*flA*1 Ap^R	This work
MK012	BL21(DE3)	Δ*crcB* p114-*flA*1 Ap^R	This work
MK013	BL21(DE3)	Δ*crcB* pET151(-) Ap^R	This work
MK014	BL21(DE3)	Δ*crcB* pSAMT Gm^R	This work
MK015	BL21(DE3)	Δ*crcB* pSAMT p*flA*1 Gm^R Ap^R	This work
MK016	BL21(DE3)	Δ*crcB* pSAMT pET151(-) Gm^R Ap^R	This work
MK017	BL21(DE3)	Δ*crcB*Δ*deoD*	This work
MK018	BL21(DE3)	Δ*crcB*Δ*deoD* pSAMT Gm^R	This work
MK019	BL21(DE3)	Δ*crcB*Δ*deoD* pSAMT p*flA*1 Gm^R Ap^R	This work
MK020	BL21(DE3)	Δ*crcB*Δ*deoD* pSAMT pET151(-) Gm^R Ap^R	This work
MK021	BL21(DE3)	Δ*deoD*	This work

Table 4.8: List of strains used in this study (continued).

Chapter 5

The effects of fluorinase expression in *E. coli*

This chapter consists of growth curve experiments showing how expression of the fluorinase affects growth of several engineered *E. coli* strains. The growth curves will be fitted using *GP* (see chapter 3). In several circumstances, Analysis Of Variance (ANOVA) was performed to show statistical significance of the results. It should be noted that the type of media and conditions used for extrapolation of growth characteristics from *GP*s, do not affect the usability of the method, since, as mentioned in section 3.1.1, the trend of growth remains sigmoidal independently of the growth medium used, as long as it is limited (not replenished).

Expression of the fluorinase in *E. coli* has been frequently performed in a number of studies, mostly in the form of overexpression by a strong inducible promoter in order to isolate ample quantities of the protein for other experiments such as *in vitro* assays for the determination of enzyme kinetics, crystal structure or others (see section 2.3). However, to the best of our knowledge, no studies that describe the *in vivo* functionality (or absence thereof) in *E. coli* or reveal

any effects that the expression exerts in the viability of the cell, apart from the fact that the protein could be readily purified. The closest match to *in vivo* fluorination has been reported in a DE study where a cell-free extract has been used to assess the enzyme's efficiency in a "relatively" high-throughput manner [52], which also pointed the author towards deletion of *E. coli* PNP. The aforementioned study also validated the choice of *E. coli* as a host for DE experiments but it was deemed that the establishment of *in vivo* fluorination would be a further enabling step, as it would open-up several possibilities for improved variant selection apart from the slow and sample-destructive techniques that are normally employed in enzyme optimization. Once the *in vivo* production of the fluorinated product of the enzyme is established, pathway engineering can reveal routes that yield directly measurable responses, therefore accelerating the DE workflow. Additionally, direct *in vivo* fluorination in *E. coli* is itself an enabling platform for other downstream applications such as organofluorine production from bioreactors. In DE (section 2.2), the simplest and most efficient screening method is genetic selection. If the *in vivo* fluorination is assumed to take place, the mitigation of fluoride toxicity by the incorporation of fluoride anions to SAM yielding 5'-FDA was visualized as a plausible mechanism of action for the selection of the highest performing variants as exhibiting a survivability advantage thus becoming gradually dominant in a mixed population. However, in *E. coli*, as in many other industrially-utilized micro-organisms, fluoride efflux channels already alleviate its negative effects simply by pumping it out of the cell (see section 4.2). The efficiency of this mechanism is tremendous, featuring a rate of efflux of more than 30,000 fluoride anions per second [111]. In comparison to the turnover of fluorinase (0.0043 s^{-1}) the effect of the latter (if any) would be negligible as fluoride efflux constitutes a competitive process 10^7 times more efficient in terms of fluoride turnover. Thus, any identification of a positive (or otherwise) effect of the fluorinase in the presence of fluoride, would have to be made on the $\Delta crcB$ strain expressing the fluorinase gene, specifically, MK010

from table 4.8, otherwise the effect of fluorinase would be masked. The idea is that the flux of fluoride anions towards 5'-FDA would be beneficial for the cell, as it would alleviate toxicity from fluoride assuming that 5'-FDA is produced within the cell, that 5'-FDA is not toxic for the cell and that there is a mechanism that expels 5'-FDA from the cell or a process that converts 5'-FDA to another non-toxic fluorinated compound which is itself exported in order to maintain a quasi-steady state equilibrium which would result in a steady fluoride turnover rate.

5.1 Growth curves of BL21 (DE3) with fluorinase expression

5.1.1 Protein overexpression side effects in *E. coli*

The overexpression of recombinant proteins in *E. coli* or yeast is an established technique for analysis of protein structure or function either with *in vivo* or, after protein isolation and purification, *in vitro* context. Expressing a foreign protein intra-cellularly however, is always risky as it can create negative effects to host cell growth. There are many different types of metabolic burden that are usually tied to growth defects. Apart from the competition for cell resources that participate in the expression, folding, localization and degradation of proteins, which constitutes the most common reason exerting a growth disadvantage even if the protein's function is orthogonal to cell metabolism, there is a possibility of toxicity due to a particular metabolic effect of a protein. This type of toxicity is apparently commonplace as depicted in a study where absence of colonies in transformed BL21(DE3) strains by chemical transformation was observed in 38% of proteins under the control of an inducible promoter with basal activity and

96% toxicity (overexpression) with the inducer [147]. The detrimental effects which commonly appear during overexpression of recombinant proteins led to the development of techniques to overcome these effects. The derivative strains C41(DE3) and C43(DE3) are results of these endeavours. There strains were selected by random mutagenesis of BL21(DE3) genotype and the ability to grow in the presence of IPTG inducer for a plasmid controlling the expression of an otherwise toxic protein for BL21(DE3) [148]. There are other tools for mitigating potential toxicity of a recombinant gene introduced to an organism through a plasmid. The pET-151/DTOPO vector used in this study for expression of the fluorinase, harbours a lacI repressor which binds both to the T7 lacUV5 promoter controlling expression of T7 polymerase in the chromosome and the lacO operator in the plasmid forbidding T7 polymerase binding in the promoter, thus further reducing basal expression levels. Except for toxicity resulting from direct interference of the overexpressed protein in cell metabolism under normal growth conditions, there is also a possibility of an indirect effect of a heterologous protein's or pathway's products to the host metabolism. For example the biopolymer PHB, consumes large amounts of key metabolites NADPH and acetyl-CoA subsequently influencing growth [149]. In this case, the stability of genetic modifications which lead to such phenotypes is at risk especially if the source of modification is a plasmid. The reason is that the growth disadvantage of the modified genotype, can become a growth advantage when loss of function occurs randomly. The clones that harbour such a loss would very quickly dominate the culture because of the growth advantage with respect to the functional phenotype which exhibits growth inhibition due to the burdensome metabolism [150].

5.1.2 Effects of fluorinase expressing plasmid in growth of *E. coli* BL21(DE3) with *Gaussian Process* analysis

Table 4.8 shows all the strains that have been developed for this work using BL21(DE3) as a base host. Many of these strains harbour expression vectors for the *flA*1 gene based on pET-151/DTOPO backbone or pJ61002. In all cases, the transformations were successful yielding ample amounts of colonies even with the strong constitutive promoter, therefore the fluorinase gene does not appear to exert toxic effects in *E. coli* and this is consistent with the absence of such reports in relevant studies (see references from section 2.3). This observation was further tested in growth experiments for a definitive statement of the effect in growth by the presence of plasmids harbouring the *flA*1 gene in BL21(DE3). Strains BL21(DE3), MK003 and MK007 were grown to saturation in LB liquid, with or without antibiotics accordingly. They were subsequently diluted 100-fold and were put in 96-well plates for time-series OD measurements as described in section 3.3.1. Each experiment was repeated in 6-8 technical replicates (same ancestral clone, different wells) and three biological replicates (different ancestral clones). The obtained measurements were then subjected to analysis with the implemented method of *GP* regression with the use of a Gompertz prior function as described in chapter 3.

Specifically from equation 3.9 and $m(t) = G(t)$ with $G(t)$ as defined in equation 3.2 and $GP(t) = E[f^*]$ denotes the fitted curve. To obtain the derivative $GP'(t)$ for estimation of the growth rate, the quantities in LHS and RHS are differentiated:

$$GP'(t) = E[(f')^*] = \partial_1 m(T^*) + \partial_1 K(T^*, T)[K(T,T) + \sigma^2 I]^{-1}(\boldsymbol{y} - m(T)) \quad (5.1)$$

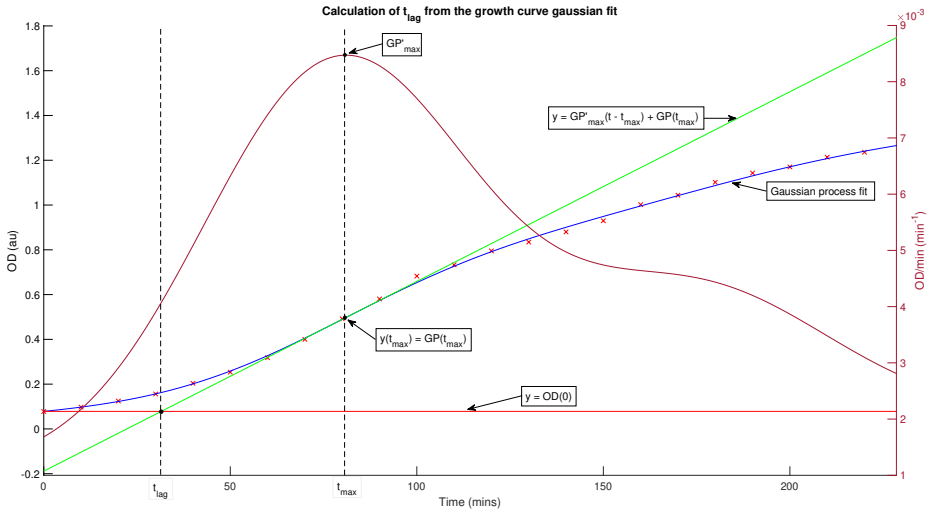


Figure 5.1: Extrapolation of lag phase from the fitted growth curve.

The derivative of the Gompertz prior function $\partial_1 m(T^*) = \frac{dG(t)}{dt}$ is obtained from equation 3.3.

Instead of looking at one quantity, namely maximum OD, the capacity of GP to fit bacterial growth curves accurately, without overfitting, offers the ability to extract more information from a curve. The extrapolated quantities chosen for comparison are the following:

- Maximum growth rate: The maximal rate of increase in OD which corresponds to the maximum rate of bacterial mass accumulation by the culture. This is obtained by the derivative of the fitted GP ($GP'(t_{max})$) from equation 5.1.

- Lag phase: The time needed by the culture in order to reach exponential growth. This is calculated as:

$$t_{lag} = t_{max} + \frac{y_0}{GP'(t_{max})} - \frac{GP(t_{max})}{GP'(t_{max})} \tag{5.2}$$

where y_0 is the starting OD of the time series, $GP'(t_{max})$ is the maximum

growth rate and $GP(t_{max})$ is the fitted OD of the culture at the time point when this rate is achieved (see figure 5.1).

- Mean OD in the stationary phase: Instead of looking at maximum OD, which is achieved only in one time point the mean OD over the stationary phase gives a more weighted value for maximum growth. The mean is obtained from the time points corresponding to the maximum value of the fitted GP curve up to the final measurement.

Calculation of these 3 attributes for strains BL21(DE3), MK003(p*flA*1), MK007(pET151(-)) yields figure 5.2. One way ANOVA shows no significant difference in lag time and maximum rate of growth, however there is significant difference in mean OD during the stationary phase between BL21(DE3) and MK007, MK003 and MK007 but not between BL21(DE3) and MK003. Apparently, with introduction of the plasmid that does not contain the fluorinase gene, there is a slight decrease in viability during the stationary phase. This is not surprising as there is some burden from the replication of the plasmid itself, the ampicillin resistance protein and the proteins that constitute the T7 expression system. The restoration of this small detriment from the introduction of the fluorinase gene, even when expressed in very low basal levels hints to an interesting effect which will be illustrated further in the next section.

5.1.3 Effects of controlled overproduction of fluorinase in *E. coli* BL21(DE3)

The experimental measurements of OD values in table 4.2, corresponding to hours after IPTG induction of strains MK003 and MK007 suggest that there is an significant increase in the OD value after IPTG induction and subsequent overexpression of the fluorinase. All experiments performed as part of this work,

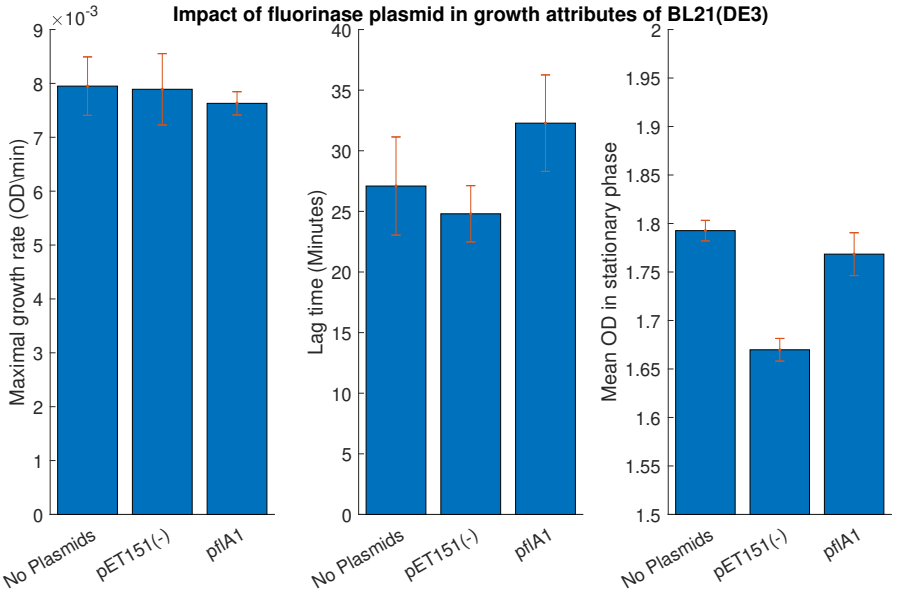
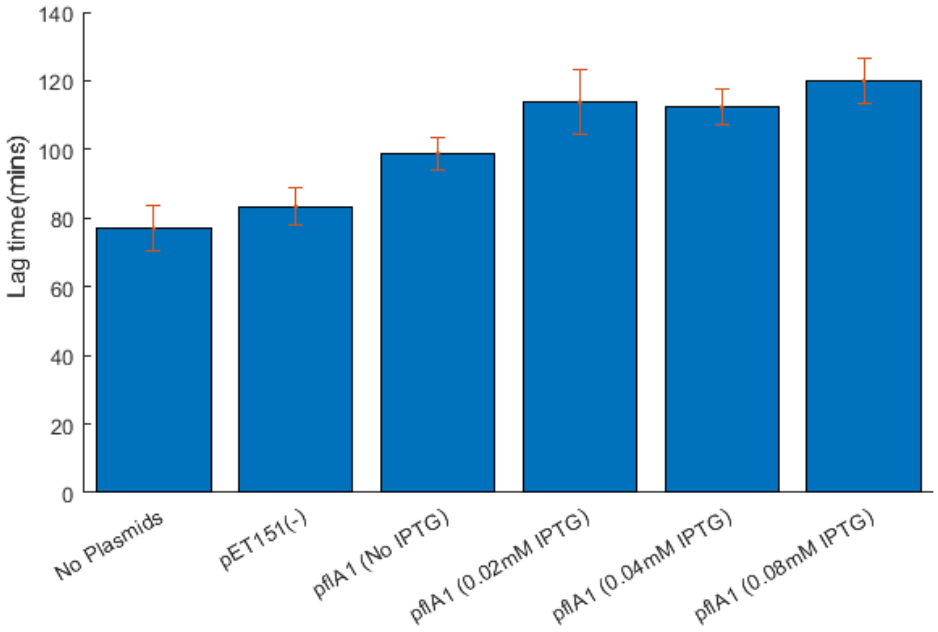


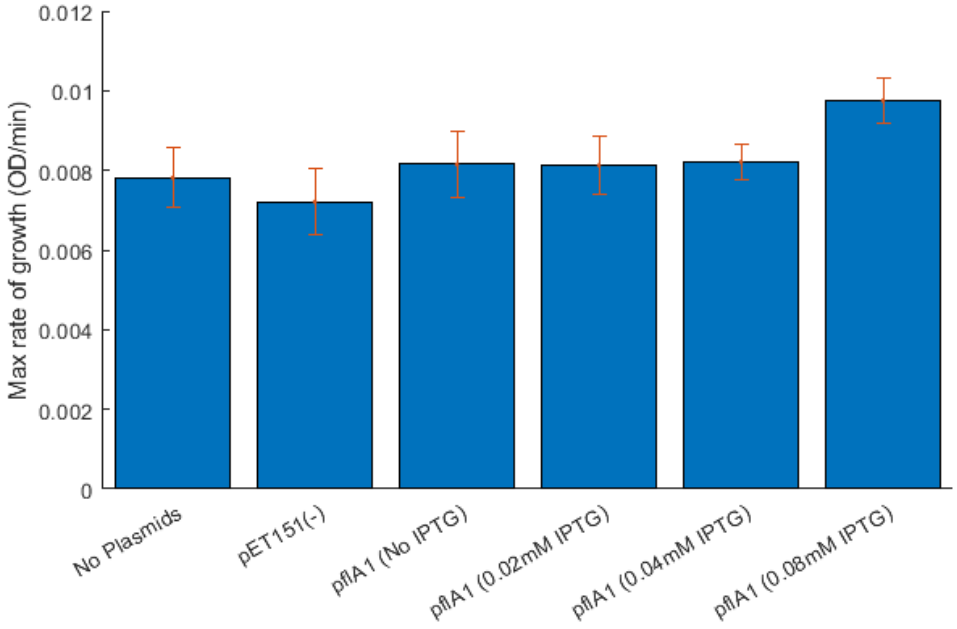
Figure 5.2: Mean values of 3 biological replicates for the attributes of growth in strains with or without plasmids. In the case of mean stationary growth there is significant difference between strains.

entailing end-point OD measurements in the liquid cultures of these strains, showed consistently higher OD in the strain bearing the p*flA*1 plasmid. To illustrate reproducibility of this result including the pure BL21(DE3) strain, an analogous experiment as in the previous section was run for the same strains but also including the conditions in which strain MK007 is grown with IPTG induction within a small range of concentrations where the expression levels vary. This was determined by another experimental dataset (obtained from A. De Las Heras) after analysis of T7 promoter activity with BL21(DE3) grown in different IPTG concentrations, with the methods developed in chapter 3.

As before, 3 different pre-cultures (biological replicates) for each strain/condition have been prepared and left to grow o/n in separate tubes. Fresh LB tubes were then inoculated with a 1:500 ratio resulting in a lower starting OD, which would enable a more pronounced determination of the differential lag time. A 96-well plate was inoculated with a number of technical replicates from each tube and

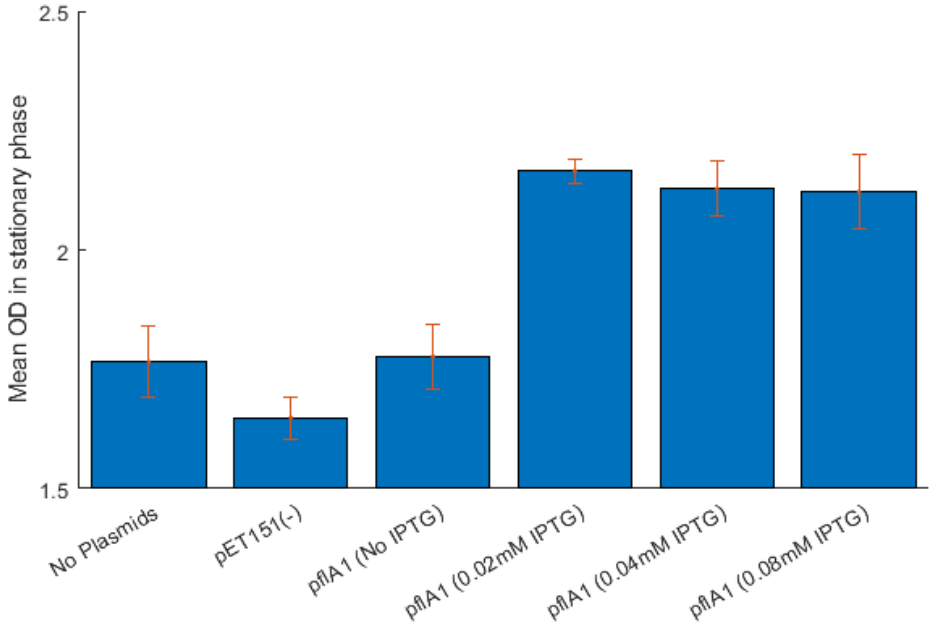


(a) Lag times

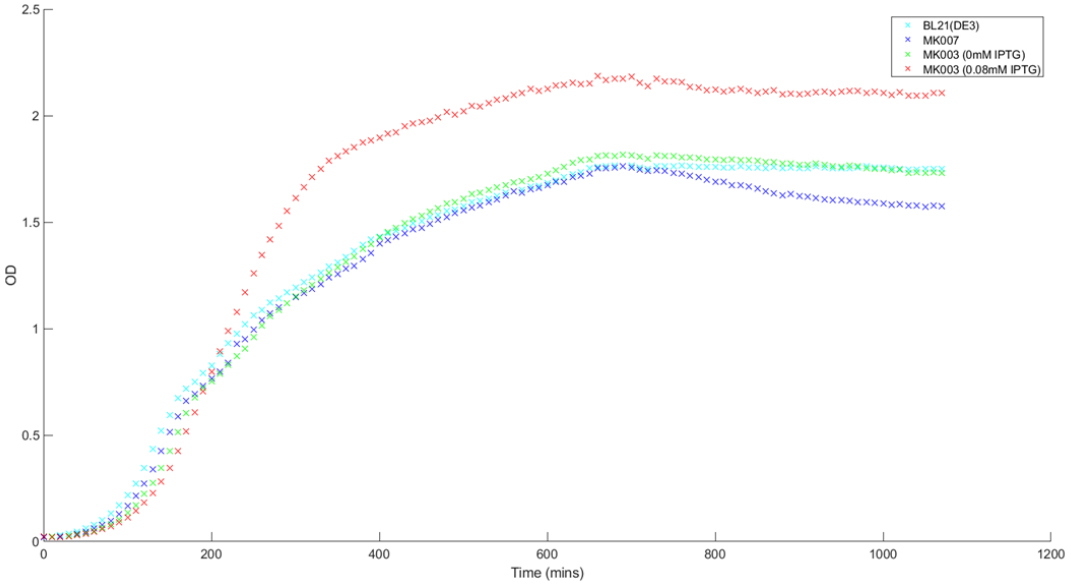


(b) Max growth rates

Figure 5.3: Lag times and maximum growth rates of strains BL21(DE3), MK007 and MK003 in varying IPTG concentrations. Lag times of strain MK003 grown in IPTG 0.02 mM or more exhibit statistically significant increase with respect to the other strains/conditions.



(a) Mean OD in stationary phase



(b) Growth curves of representative conditions

Figure 5.4: Mean stationary OD and mean growth curves of strains BL21(DE3), MK007 and MK003 in varying IPTG concentrations. The higher ODs in IPTG induced cultures is a statistically significant result.

OD measurements were performed over 18 hours with a time step of 10 minutes. The chosen IPTG concentrations where 0 mM, 0.02 mM, 0.04 mM and 0.08 mM corresponding to around 10%, 20%, 40% and 80% of maximum expression levels respectively (reached with 0.5 mM and 1 mM). Strains MK007 and BL21(DE3) were not tested in the presence of IPTG as in a recent study, BL21(DE3) growth shows an identical trend [151]. In the same study, it is also highlighted that the overexpression of a neutral control protein imposes a burden to cell growth, albeit small.

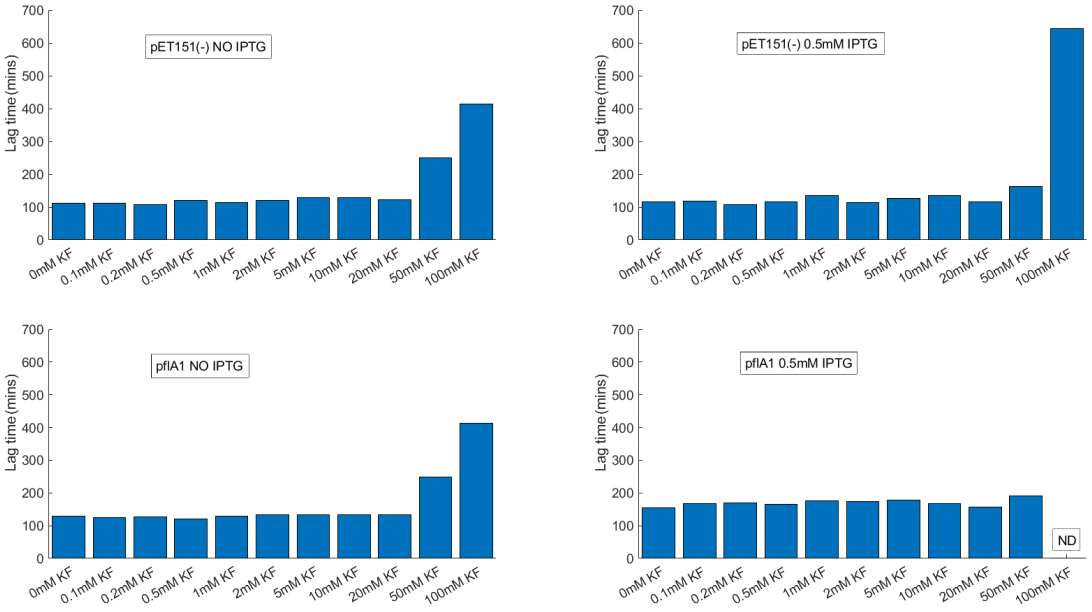
The time series obtained were subjected to the same analysis with *GP* fitting as in section 5.1.2 and the same 3 attributes were extracted. In figures 5.3 and 5.4, the means and deviations of these attributes are illustrated.

From figure 5.3a, it is evident that there is an increasing trend in the lag times with the strains harbouring the fluorinase with increasing IPTG concentrations, however between neighbouring conditions the differences are borderline or not significant. However, between strains BL21(DE3) and MK003 with 0.08 mM IPTG induction, the statistical significance in lag time is indisputable, with the latter exhibiting a higher delay in reaching exponential phase. This is a result to be expected as the induction of high level protein expression diverts a good portion of resources towards this task, thus there is congestion in usage of the same resources for cell growth.

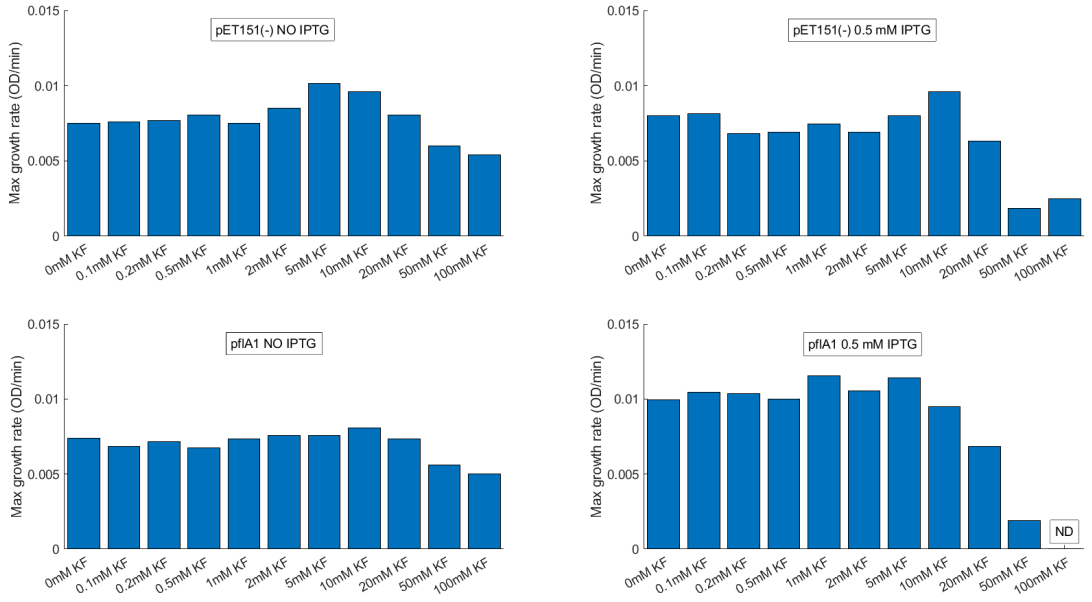
The maximum rates of growth are illustrated in figure 5.3b. The only mentionable result is that the statistically significant difference, based on ANOVA, in the maximum growth rate achieved between MK007 with IPTG of 0.08 mM and all other strains/conditions. The remainder cultures of MK007 with IPTG concentrations 0, 0.02, 0.04 mM do not show significant difference between them or with strain BL21(DE3) however they show significant difference with MK003. The most profound result is seen in the mean OD during stationary phase where

there is a definite increase in OD observed with strain MK003 grown in all non-zero concentrations of IPTG. Specifically, even in the low-level overexpression (20% according to the calibration experimental data from A. de las Heras), the highest OD is reached as shown in figure 5.4a. The difference of means with strain BL21(DE3) grown in pure LB, is 23%. All non-zero IPTG concentrations reach OD values very close to each other (no statistical significance). This is a stunning result, in the sense that, not only there seems to be a positive effect in the *E. coli* cells with overexpressed fluorinase, since they outgrow simple BL21(DE3) cells grown in rich medium, but this effect takes place in low and persists in high overexpression levels. In order to identify the source of this effect, possible mechanisms of action where envisaged, such as the dechlorination of LB, which contains NaCl, by the duality of function of the fluorinase as a chlorinase [4]. This was however not the case, as the same differences in OD were observed with media prepared in the absence of NaCl (data not shown). Another plausible scenario is that the SAM binding of the enzyme could create a transient imbalance in intracellular SAM, as recognised by SAM riboswitches, which could subsequently lead to higher intracellular SAM production, in turn stimulating further growth. In any case, not excluding the possibility of a non-specific reaction of the fluorinase, the reason for this positive growth effect is most probably tied to substrate binding in the enzyme.

The significance of the above observation extends in the usability of this enzyme within *in vivo* reactions in *E. coli*. As mentioned in section 5.1.1, in most procedures involving recombinant protein expression in a foreign host, the risk of gene instability is imminent with the extra metabolic burden imposed by this protein. The positive effect of fluorinase expression in *E. coli* not only eliminates the need for a gene stabilization strategy, but also provides a metabolic surplus for the potential encumbrance from the participation of this enzyme in recombinant pathways.



(a) Lag times with increasing fluoride



(b) Max growth rates with increasing fluoride

Figure 5.5: Lag times and max growth rates of strains MK003 and MK007 with/without IPTG. In IPTG-induced strain MK003 with 100mM KF, lag time and max growth rate cannot be determined as the measurements finished before entering the exponential phase.

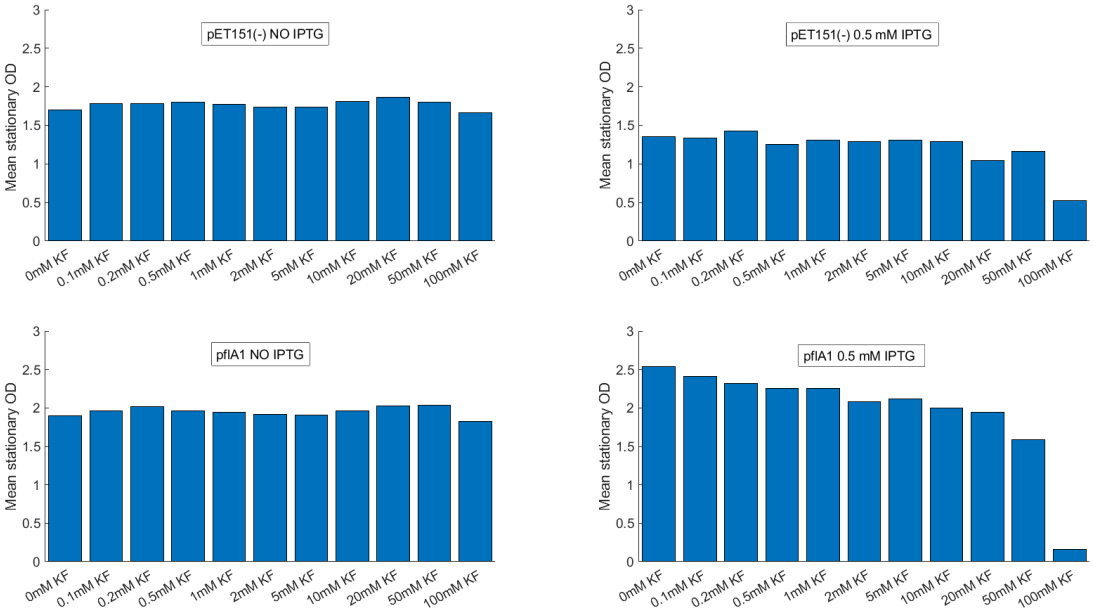


Figure 5.6: Mean stationary OD of strains MK003 and MK007 with/without IPTG.

5.1.4 Fluoride dependent growth of BL21(DE3)

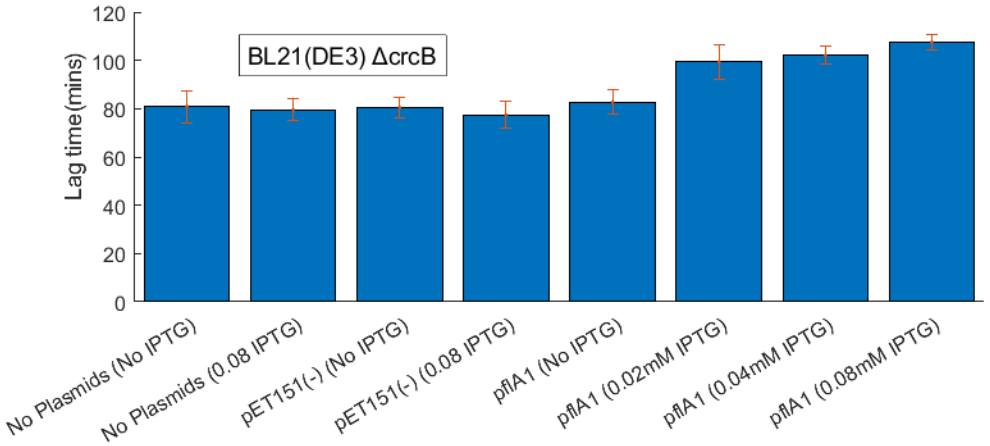
Strains MK003 and MK007 were grown as before, but including various concentrations of potassium fluoride in order to quantify its effect on the growth attributes of these strains in basal expression levels or in maximum overexpression. To this end, the strains were grown both in 0 and 0.5 mM of IPTG and the increments in concentrations of fluoride were performed in a logarithmic manner. After *GP* analysis and extraction of attributes, results are collectively illustrated in figures 5.5 and 5.6.

A few interesting observations can be made with respect to these measurements. First of all, in fluoride concentrations up to 20 mM there is little to negligible effect in all aspects of growth without IPTG addition. This is to be expected, as in these concentrations, the functional fluoride efflux channel keeps the intracellular concentration more than 100-fold less of the extracellular (see figure 4.7). In 50 and 100 mM, as the detrimental effect of fluoride sets in, there is a large increase in lag times, and a measurable decrease of max growth rates, while stationary

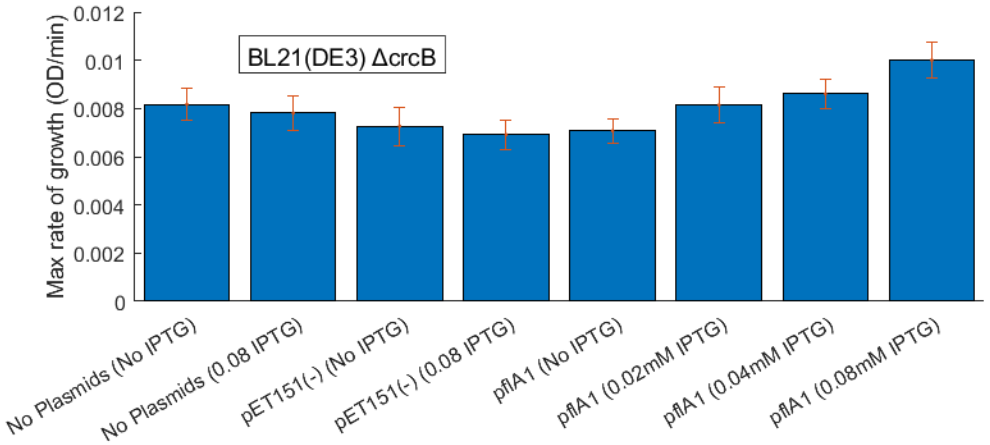
OD remains at similar levels. Evidently, at concentrations when fluoride begins to exhibit a deleterious effect, it impacts mostly on lag time, with max growth rates and stationary OD's being affected considerably less. In IPTG-induced growth, lag times of strain MK007 remain roughly the same up to 50 mM and in similar ranges with the uninduced strains. On the other hand, the induced MK003 strain (expressing the fluorinase), shows higher lag times than the rest of the strains but also reaches higher stationary ODs, and maximum growth rates, illustrating the positive effect of fluorinase overexpression on growth, in accordance with the experiments of the previous sections. Interestingly, there is a negative effect on stationary OD of the IPTG-induced MK007 in comparison to the uninduced state. Effects on lag times and stationary ODs in IPTG induced strains take place in 100 mM of fluoride and in the case of fluorinase overexpression they are much more pronounced indicating a synergistic detrimental effect of fluoride and fluorinase. However, because of the competitive nature of the CrcB channel expelling fluoride, it is difficult to draw conclusions about the effect of the fluorinase in different fluoride concentrations, therefore an analogous growth experiment was conducted but with strains MK010 and MK013 instead where the CrcB channel is knocked-out and the extracellular fluoride concentration is more representative of the intracellular (see next section).

5.2 Growth curves of the Δ*crcB* and Δ*crcB*Δ*deoD* strains with fluorinase expression

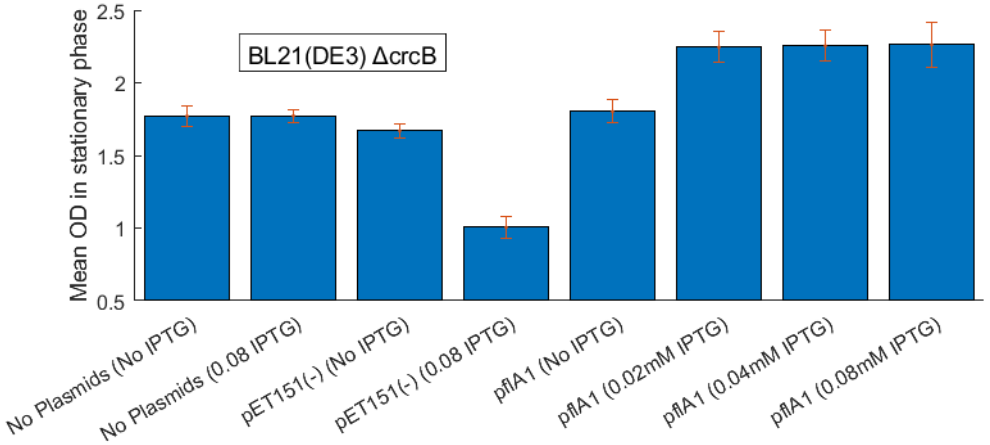
Argumentation for the deletion of the fluoride efflux channel and the *E. coli* PNP genes (*crcB* and *deoD*) was presented in the previous chapter. A gene deletion, even when its function has been determined, as is the case with *crcB* and *deoD*, cannot rule out secondary effects in combination with other conditions. Therefore, while it is expected that there would be similar behaviour in the conditions tested



(a) Lag times



(b) Growth rates



(c) Mean stationary OD

Figure 5.7: Mean stationary OD and mean growth curves of strains MK009, MK013 and MK010 in varying IPTG concentrations.

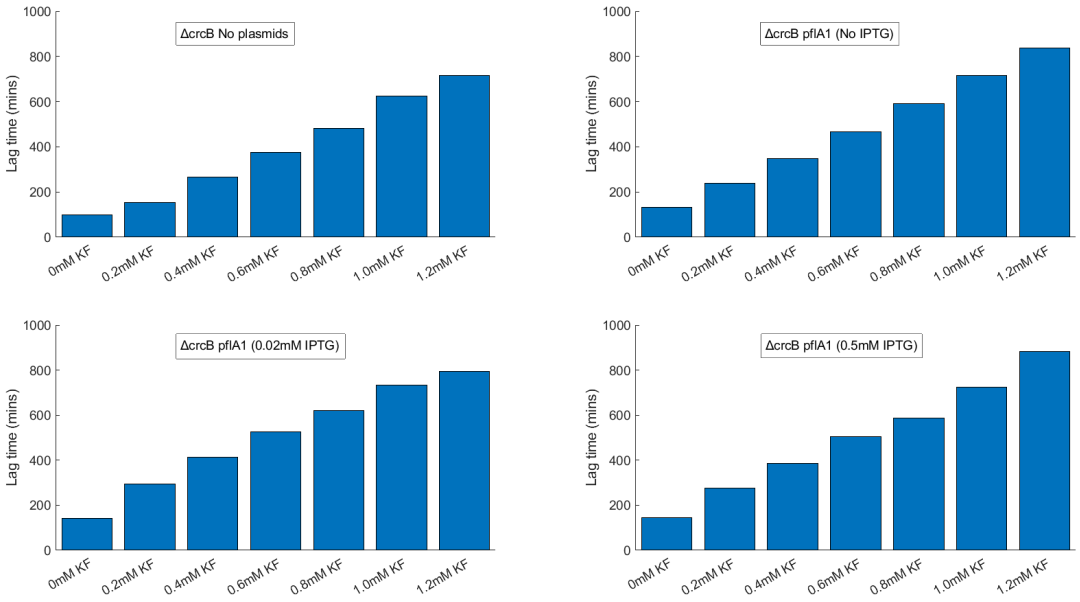
with BL21(DE3), the question rises whether the positive effects from fluorinase expression, extend to strains accommodating these deletions.

5.2.1 Growth of Δ*crcB* and Δ*crcB*Δ*deoD* strains

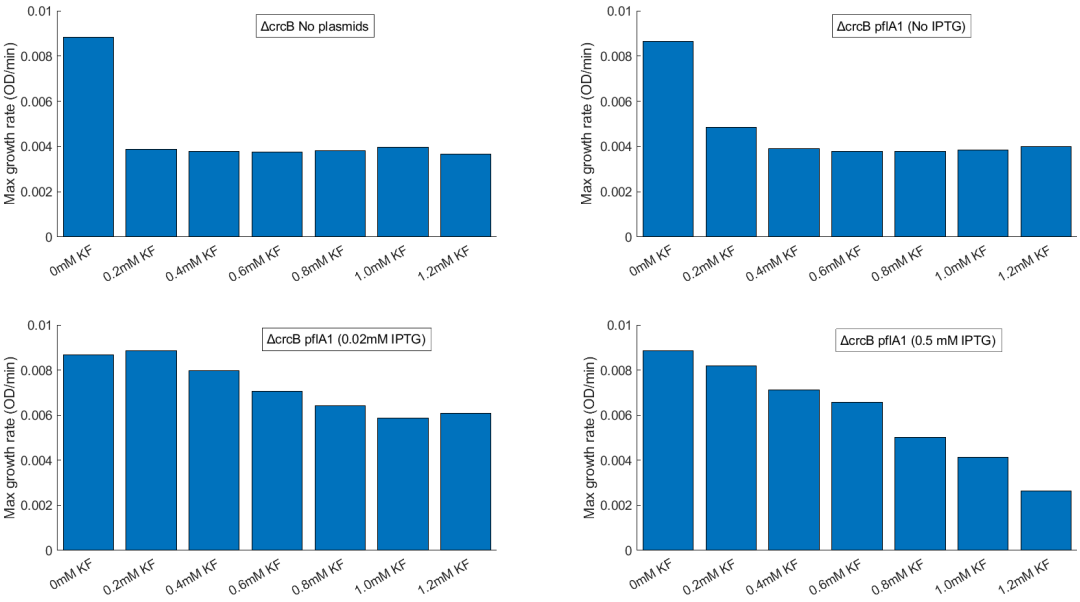
First of all, the single deletion strain with *crcB* KO was tested. Specifically, strains MK009 (Δ*crcB* no plasmids), MK013 (Δ*crcB* pET151(-)) and MK010 (Δ*crcB* p*flA*1) were grown in a 96-well plate and an OD time-series of 10 minute intervals over 18 hours was obtained as previously described. This time, in order to test the effect of IPTG without the p*flA*1 plasmid, strains MK009 and MK013 were grown in 0.08 mM IPTG in addition to no IPTG. The results are collectively illustrated for all measures extracted from growth curves in figure 5.7.

In comparison to BL21(DE3), there is similar behaviour in these strains with and without plasmids. In general, growth attributes do not show a decline as a result of this deletion. Similarly, lag times show an increasing behaviour with higher levels of expression, as do max rates of growth, but most importantly the positive effect of fluorinase from low to high levels of expression persists in this strain, leading to higher accumulation of bacterial mass. Another interesting observation is the behaviour of MK013 under induction of IPTG which shows a considerable decrease in stationary OD. This is not an effect of the IPTG itself, as strain MK009 with induction does not show a similar behaviour, therefore it is probably related to a polar overexpression effect in plasmid pET151(-) after removal of the fluorinase gene.

Growth in strains bearing the double deletion genotype exhibit similar effects, however, compared to the Δ*crcB* or BL21(DE3) phenotypes, growth of these strains appear somewhat decreased. These results are not shown as this particular deletion does not ultimately affect the ability of *E. coli* to produce fluorometabolites (see section 5.4).

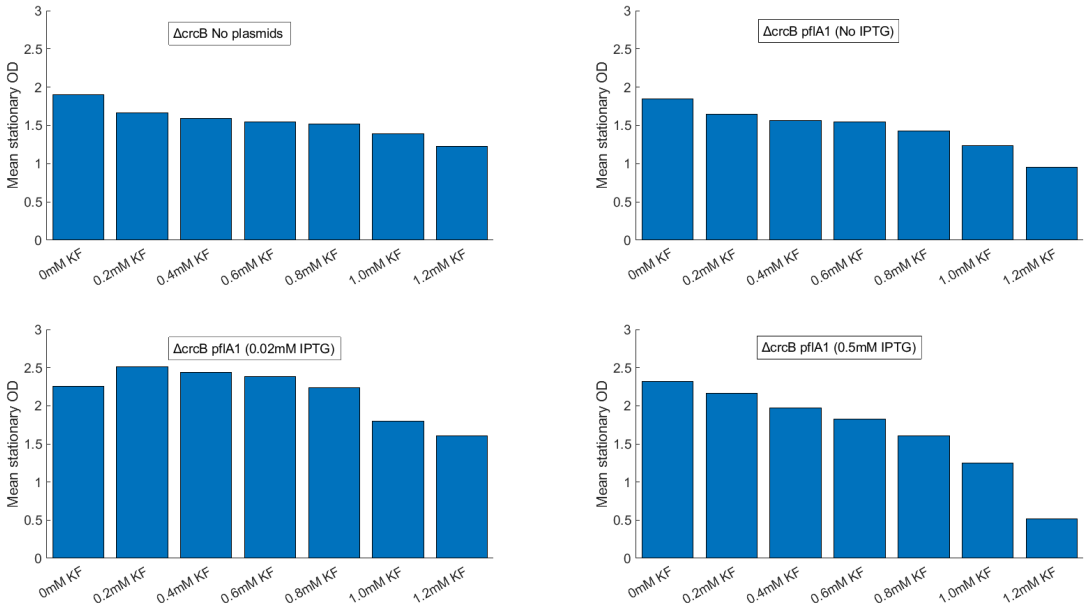


(a) Lag times with increasing fluoride

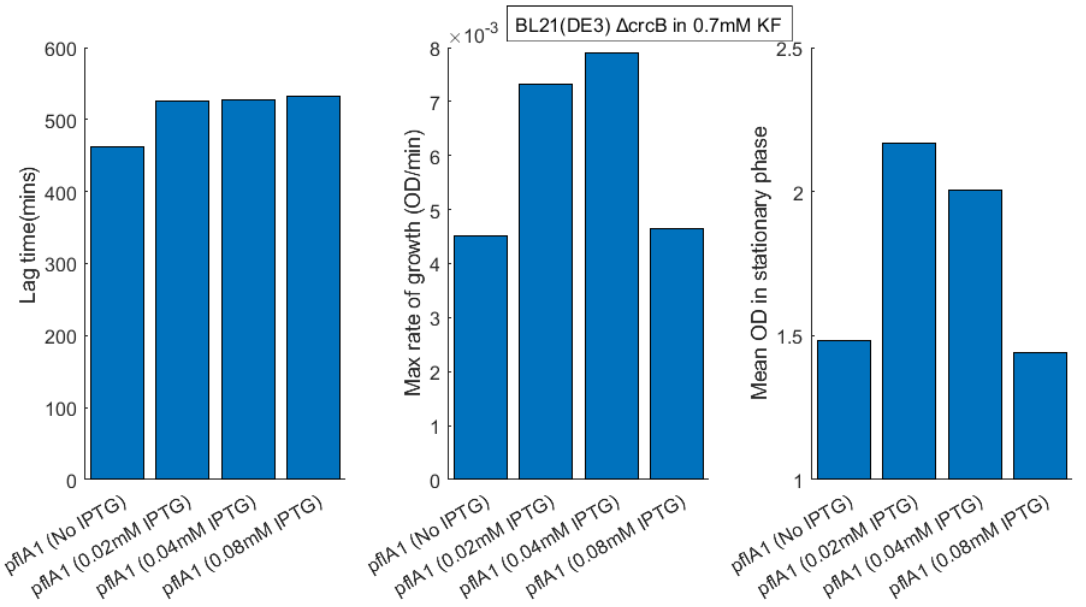


(b) Max growth rates with increasing fluoride

Figure 5.8: Lag times and max growth rates of strains MK009 and MK010 in various IPTG concentrations.



(a) Mean stationary OD with increasing fluoride



(b) Growth attributes in 0.7 mM fluoride and varying expression levels

Figure 5.9: Mean stationary OD of strains MK009 and MK010 and growth attributes of strain MK010 with fluoride 0.7 mM and in various IPTG concentrations.

5.2.2 Fluoride dependent growth of $\Delta crcB$ and fluorinase expression

As previously described by a number of studies (see 4.2.3) and confirmed by phenotype validation, strains with a *crcB* deletion are expected to show a reduction in growth with added fluoride in molarities starting from micromolar up to 2 mM. With the fluoride efflux channel deleted, the fluoride molarity of the growth medium is more descriptive of the intracellular molarity and can be easily fine-tuned. To this end, a range of linearly increasing fluoride molarities have been tested, from 0 mM to 1.2 mM with 0.2 mM increments. In this instance, only strains MK009 and MK010 have been tested as the unexpected effect of plasmid pET151(-) in growth of MK013 under IPTG induction, illustrated in the previous section, disqualified this strain as a reliable negative control for growth curve comparison. It will however be used as a negative control for 5'-FDA production (see section 5.4). Three values of IPTG molarities have been tested in strain MK010 corresponding to basal, low-level expression and overexpression (0, 0.02 and 0.5 mM).

The value of 0.02 mM has been chosen from an optimization experiment whose results are shown in figure 5.9b. The latter has been performed with a growth-reducing concentration of 0.7 mM fluoride and in varying expression levels, in order to identify effects in growth from the action of different amounts of intracellular fluorinase. The difference in lag times is consistent with this seen in previous experiments and has to do with the devotion of resources in overexpression. There is a clear difference both in max growth rates and in mean stationary ODs in different expression levels following the same pattern. In comparison to basal expression, there is a large increase in growth rate with low level (0.02 mM IPTG) which shows further increase in mid-level (0.04 mM), while in high level there is a decline back to the same value as the basal level.

With stationary OD the same effect is observed but the highest value is achieved with low level of expression. This is an insightful result which shows that indeed the presence of intracellular fluorinase in low to mid-levels does exert a positive effect in growth in the presence of a fluoride concentration that negatively impacts growth. It is also noteworthy that this effect reverts in higher levels of expression.

Back to the experiment with increasing fluoride, the growth attributes have been extracted by *GP* analysis and the results are shown in figures 5.8 and 5.9a. In all cases, lag times follow an almost linear increasing pattern with fluoride concentrations. There is no difference in this trend with different expression levels. Growth rates on the other hand show a different behaviour. Both in strain MK009 and MK010 with basal expression, there is an abrupt drop even with 0.2 mM KF, to less than half of the max growth rate without fluoride and it is maintained in up to 1.2 mM. However, in MK010 with low overexpression, in 0.2 mM KF there is even an increase in growth rate which drops in higher KF concentrations but is always at 50% more than the one seen in strains with no fluorinase or basal expression thereof. In maximum overexpression, growth rate drops gradually but faster than with lower expression and is maintained in higher values up to 1.2 mM where it drops to even lower levels than in basal or no expression. The stationary OD also exhibit a profound effect. In basal or no expression there is a similar decreasing pattern starting with values lower than 2 (OD au). In 0.02 mM IPTG induction, apart from the increase in OD due to the positive effect of the fluorinase itself in growth, as shown in previous sections, there is even an increase in stationary OD with 0.2 mM KF which slowly drops up to 0.8 mM at which point it becomes roughly equal to OD without added KF, still above the value of 2. On the other hand, with overexpression there is a decrease starting from 0.2 mM KF ultimately reaching a value lower than any other condition in 1.2 mM KF. This is an example where the metabolic surplus created from the positive action of fluorinase in growth takes effect and maintains the level of stationary OD at values that are higher than in the cases where fluorinase expression is basal

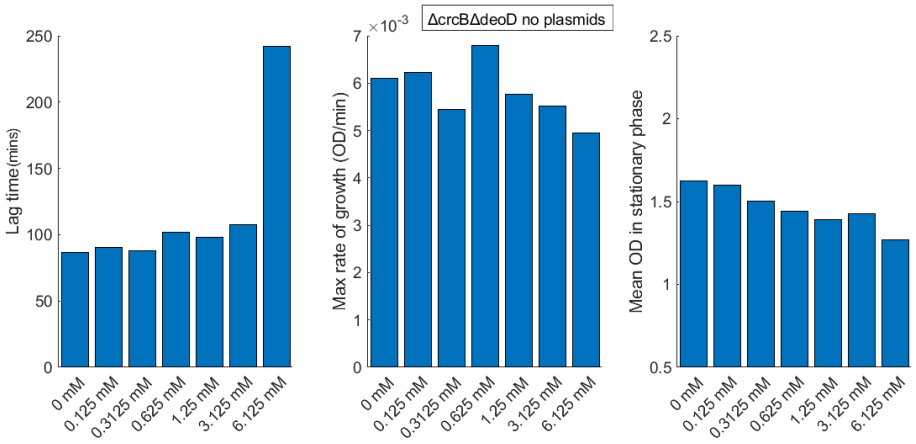


Figure 5.10: Increasing SAM effect on *E. coli* without a SAM transporter (MK017)

and below, even in concentrations of fluoride that are proven to be detrimental in the growth of *crcB* KO strains.

5.3 Effects of increasing SAM concentration in growth

The last modification towards the creation of a fluorination-enabled *E. coli* host, described in the previous chapter, is the expression of SAM transporter from *R. prowazekii*. The kinetics of this membrane protein have been characterized, and interestingly they differ between its native host and when expressed heterologously in *E. coli* [124]. According to this study, the process of SAM uptake reaches a plateau in *R. prowazekii* within 8 minutes of exposure to 10 μM extracellular SAM and it still achieves an in:out ratio of more than 10. In *E. coli* the uptake process maintains a steady rate for at least 20 minutes. It has not been yet determined if a maximum is reached after 20 minutes. Moreover, it has been shown from the same group (Wood) that the *metK* deletion strain complemented by the SAM transporter (MOB1490) shows optimal growth in 17.5 μM of added SAM [142] and the intracellular SAM concentration during exponential phase of growth has

been measured to be 180 μM [123]. From these data it is safe to assume that in *E. coli* the inside to outside ratio is more than 10 at least in low molarities. The advent of SAM transporter expression in *E. coli* is the ability to vary intracellular SAM levels at will, since SAM is the key substrate in the fluorinase reaction and since intracellular presence of fluoride has been established from the *crcB* KO. It is of interest therefore, to identify the effects of increasing SAM molarity in *E. coli* growth, in milimolar levels (100 times more than has been tested before) with or without a functional transporter or *metK* gene.

5.3.1 SAM chloride dihydrochloride effect in growth of *E. coli* without SAM transporter

SAM, apart from being a cation, it is also highly unstable in room temperature or higher, and spontaneously degrades to 5-MTA and homoserine lactone by cleavage [152]. Therefore, it is normally obtained in a chemical form where stability is somewhat increased (SAM chloride dihydrochloride - SIGMA). However, dihydrochloride will alter the pH of the growth medium, and during a full-day incubation, the degradation products will constitute a considerable percentile of the total substrate added and may also exert secondary effects to bacterial growth. Therefore, as a control experiment, strain MK017 (double KO - no plasmids) was grown in the presence of varying SAM concentrations (0 to 6.25 mM) in order to identify secondary effects of SAM in this chemical form that are irrelevant to transporter functionality. Results for these growth experiments are shown in figure 5.10.

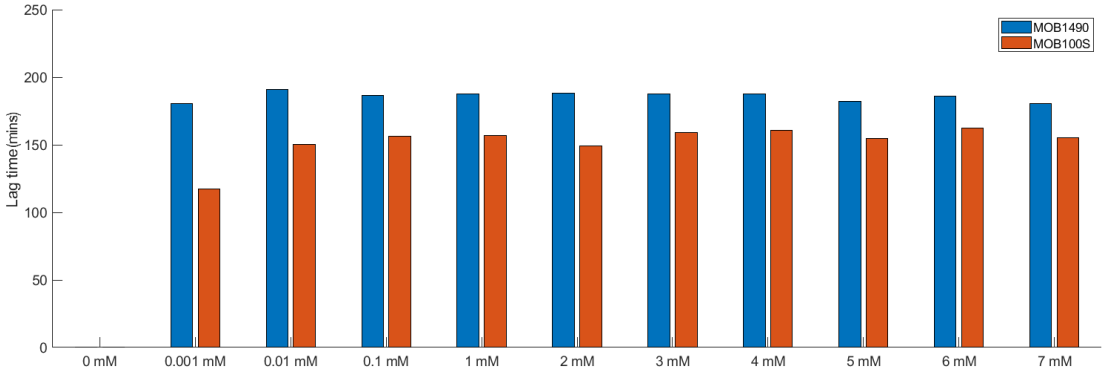
Lag times in strain MK017, exhibit an increasing behaviour with higher values of SAM but the differences are minor and within two measurement time steps (20 minutes). This changes in the last increment from 3.12 mM SAM to 6.25 mM where lag time is more than doubled from this change. Clearly, there is a

detrimental effect of high molarity SAM in the lag phase of growth which is not related to SAM uptake. Regarding max growth rate and mean stationary OD there is a decreasing trend in both, with minor differences in between changes, however in the extremes of no SAM addition and mM levels of SAM they add up to a considerable difference. Again this shows that there is a secondary cause of altering growth attributes in higher molarities of SAM which is most probably connected to the change of pH towards acidic with the addition of hydrochloride molecules, without excluding the possibility of the degradation products from SAM cleavage inhibiting growth.

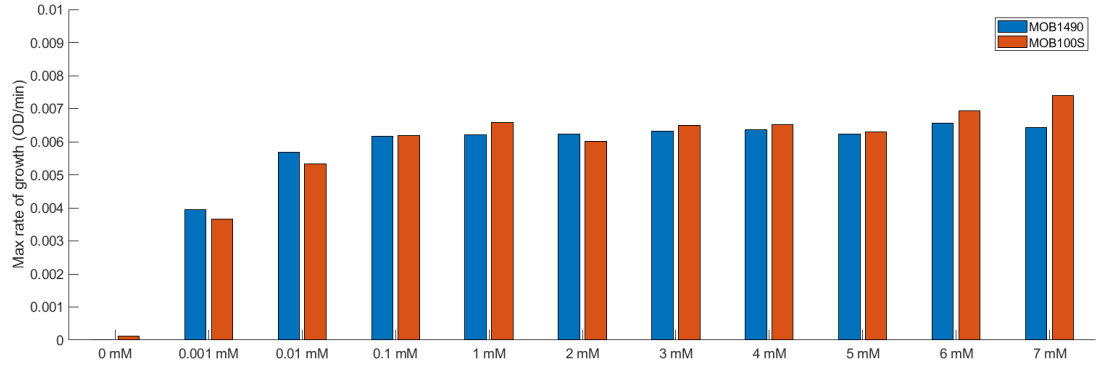
5.3.2 Growth of *metK* deletion strains with SAM transporter

In this growth experiment, SAM concentration was pumped-up to millimolar levels in order to identify the effects of high concentrations in the *metK* deletion strain bearing the SAM transporter. Both strains carrying constructs with the *samT* gene were tested, namely MOB1490, the strain originally used for the identification of transporter functionality, and strain MOB100S, which was derived from MOB1490 with a plasmid exchange method and carries p100-SAMT (see tables 4.6 and 4.7). The results after *GP* analysis are shown in figure 5.11.

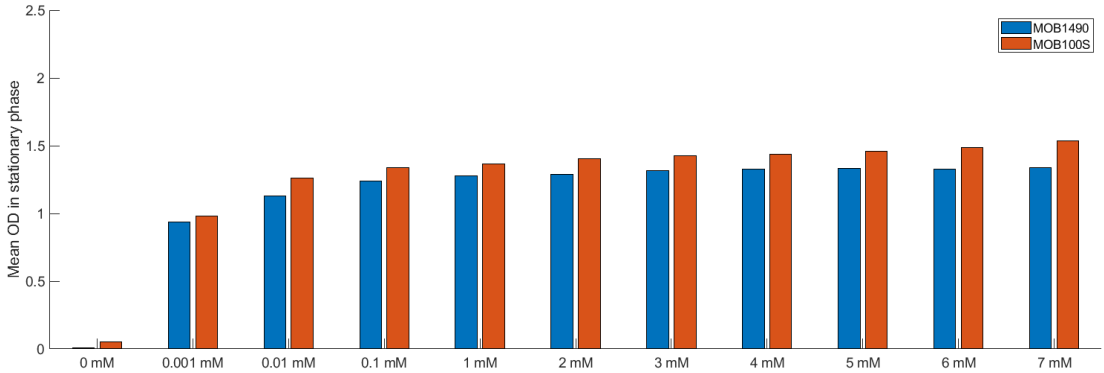
There are measurable differences between strains in two aspects of growth, lag time and mean stationary OD, both favourable for strain MOB100S. These strains differ in two ways. First of all, in MOB100S the gene was synthesized in a codon-optimized manner while in MOB1490 the original sequence from *R. prowazekii* was used. Second, in MOB100S, a different backbone plasmid was used and the promoter fusion was J23100 resulting in strong constitutive expression, while in MOB1490 the backbone used was pSMART and the gene expression was constitutive but with the native promoter.



(a) Lag times



(b) Growth rates



(c) Mean stationary OD

Figure 5.11: Growth attributes of strains MOB1490 and MOB100S in various concentrations of SAM chloride dihydrochloride.

Some observations are noteworthy. In figure 5.11, 0 mM is actually $3.5x10^{-5}$ mM because there was a 1000-fold dilution in LB from a pre-culture containing 0.035 mM SAM. Interestingly, there is zero growth in this concentration of SAM in MOB1490 but there is a slight growth curve in MOB100S. This is evidence that higher expression levels of this active one-way transporter, promotes growth in extremely low concentrations of SAM. In 1 μM there is not a significant difference in stationary OD between strains but there is around 50% difference in lag time with MOB100S being faster in growing showing again a more efficient uptake process. In higher concentrations, the differences in lag times become lower but are always in favour of MOB100S, while in mean stationary ODs the differences increase and equilibrate at a level of 15% higher in MOB100S, in the highest concentration of SAM tested (7 mM). Interestingly, in this concentration, max growth rate is also higher in MOB100S. Compared to growth of $\Delta crcB\Delta deoD$ in various concentrations, the negative effect seen in 6.25 mM does not seem to appear in neither of MOB1490 or MOB100S. A possible explanation is that excessive SAM promotes growth and counterbalances the inhibition. Evidently, the higher amounts of SAM transporter being expressed in MOB100S enhance the ability of cells to uptake SAM from the medium.

5.4 *In vivo* fluorination

The growth attributes in several combinations of strains/conditions have been illustrated in the previous sections, indicating some synergistic effects from the levels of fluorinase expression, SAM transport functionality and fluoride concentration, however the most important question is whether the reaction is happening inside the living and dividing modified *E. coli* host and which fluorometabolites are produced. In the following sections, the validation of *in vivo* direct fluorination in *E. coli* will be illustrated. Specifically, a range of

chemical analysis methods were utilized for the identification of fluorometabolites, namely High Performance Liquid Chromatography (HPLC), Fluorine-19 Nuclear Magnetic Resonance (^{19}F-NMR) coupled and decoupled, Proton nuclear magnetic resonance (^{1}H-NMR) and High Resolution Mass Spectrometry (HRMS). In particular, ^{19}F-NMR was quite useful in this experimental context, as it only detects molecules with fluorine atoms embedded.

This series of experiments presented in the sections below were envisaged by the author but in their details, they were jointly designed with collaborators from the chemistry department of the University of St. Andrews Dr. Phillip Lowe and Professor David O'Hagan. The incubations of *E. coli* for assaying 5'-FDA productivity were performed mostly by the author but with a significant contribution from Mr Liam Davidson-Gates. Chemical analysis of samples derived from incubations were performed, with all methods listed in the previous paragraph and presented throughout this section, by Dr Phillip Lowe.

5.4.1 Experimental conditions

Strains MK015 and MK019 were picked as most promising for demonstrating the capacity to produce fluorometabolites *in vivo*. It also seemed possible that any potential degradation from *E. coli* PNP which is not deleted in MK015 would elaborate additional fluorometabolites in the cytoplasm *E. coli* and this would also constitute an important finding.

Incubation conditions

Colonies of strains that have been previously validated with PCR, chosen for each separate condition/strain combination, were picked from agar plates and used for starting fresh liquid LB cultures. These were grown to saturation and subsequently diluted 100-fold in 250 mM Erlenmeyer flasks containing 100 mL of

LB. SAM chloride dihydrochloride was also added at this point according to the conditions of the particular experiment and the flasks were put at 37 ^{o}C in a rotary shaking incubator (170 rpm). When each culture reached mid-log phase (OD of 0.5-0.7, approximately 2 hours) it was induced with 0.2 mM of IPTG. According to the calibration experiment (see section 5.1.3), this molarity of IPTG results in protein expression just below the maximum levels. After 5 hours, when according to the pilot expression experiment (see 4.1.3), maximum fluorinase expression is achieved, potassium fluoride was added at a final concentration determined by the specific conditions of each incubation. Cultures are subsequently incubated for 24 more hours to allow for the fluorination reaction to equilibrate and then centrifuged in 3,000 rpm for 20 minutes. This low speed ensures minimal breakage of cells from stress and unintentional release of 5'-FDA in the S/N. Cell pellets and S/Ns were kept in -20 ^{o}C before chemical analysis.

Denaturing and lysis of cells

For analysis of the intracellular metabolites, the cells were denatured/lysed as described in [153]. Pellets acquired after incubation were added 70% ethanol in water solution pre-warmed at 70 ^{o}C and vortexed for 30 seconds. Then they were put in 95 ^{o}C and left to boil for 5 minutes followed by 5 minutes in 4 ^{o}C. The denatured cells were then subjected to a 5 minute centrifuge cycle in 14,000 rpm to separate cell debris from the lysate.

5.4.2 Establishment of the products of fluorination in the cytoplasm

Two pairs of concentrations of the reactants were tested for each of strains MK015 and MK019 and their negative control strains (MK016 and MK020), 2 mM/50 μM and 15 mM/250 μM of KF/SAM respectively, adding up to a total of 8 samples.

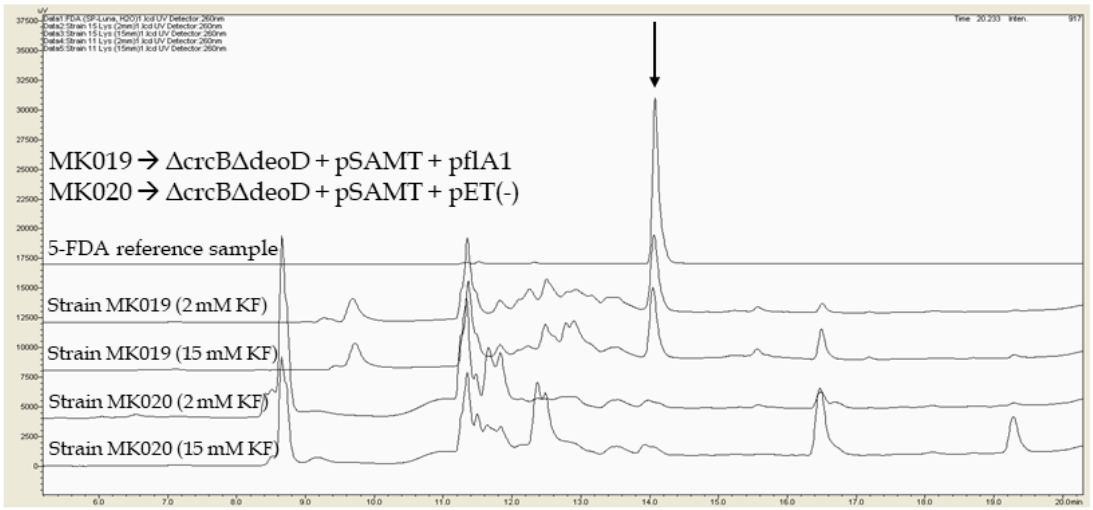
The goal of this experiment was to detect fluorometabolites in *E. coli* and identify differences between strains/conditions. The first step was HPLC analysis in both the S/N and lysate corresponding to each sample. Results are shown in figure 5.12.

In the HPLC trace of the lysates (fig. 5.12a), there is a distinct difference in the intensity of the peak corresponding to 5'-FDA in strain MK019 compared to the peak appearing in strain MK020 in both conditions. This is an encouraging observation but not definitive of the presence of 5'-FDA. Moreover, there does not seem to be a difference in the intensity of peaks between conditions 15 mM/250 μM and 2 mM/50 μM of KF/SAM. This could be attributed to a saturation of the process intracellularly or the debilitation of cells and subsequently of proteins, due to the extra metabolic burden imposed by such a high molarity of fluoride. According to [113], an intracellular concentration of 15 mM is not enough to kill the cells after >24h of exposure to fluoride, although it halts growth and reverts the cells back to lag phase, contrary to lower cytoplasmic fluoride which does inhibit growth but cells remain in exponential phase as shown when growth is resumed in the absence of fluoride. High fluoride exposure must therefore induce a stress response to the cells probably leading to a decrease in protein levels as they struggle to maintain a minimal metabolic demand. Strains MK015 and MK016 (functional CrcB) were also analysed and exhibit the same behaviour.

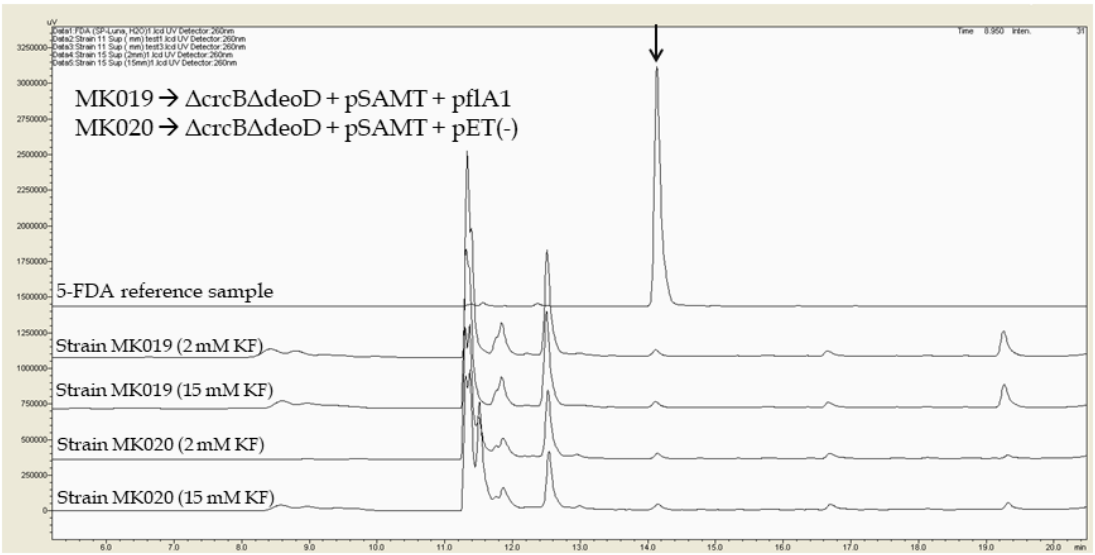
The HPLC traces of the S/Ns exhibit a small peak in the same retention time, however it is present in the negative control strain MK020 as well, therefore it cannot be attributed to extracellular 5'-FDA. Apparently, this peak belongs to compound(s) that make up LB with similar retention times as the trace of plain LB also exhibits this peak (data not shown).

Validation with ^{1}H-NMR, ^{19}F-NMR and HRMS

Strain MK015 lysate was subjected to ^{19}F-NMR analyis and results are shown



(a) Lysate



(b) S/N

Figure 5.12: HPLC traces of the cell lysates and S/Ns of strains MK019 and MK020 after 24h of incubation with 2 mM or 15 mM KF and 50 µM or 250 µM of SAM respectively. The HPLC trace of pure 5'-FDA was also obtained as a reference under the same conditions and its retention time is 14 min.

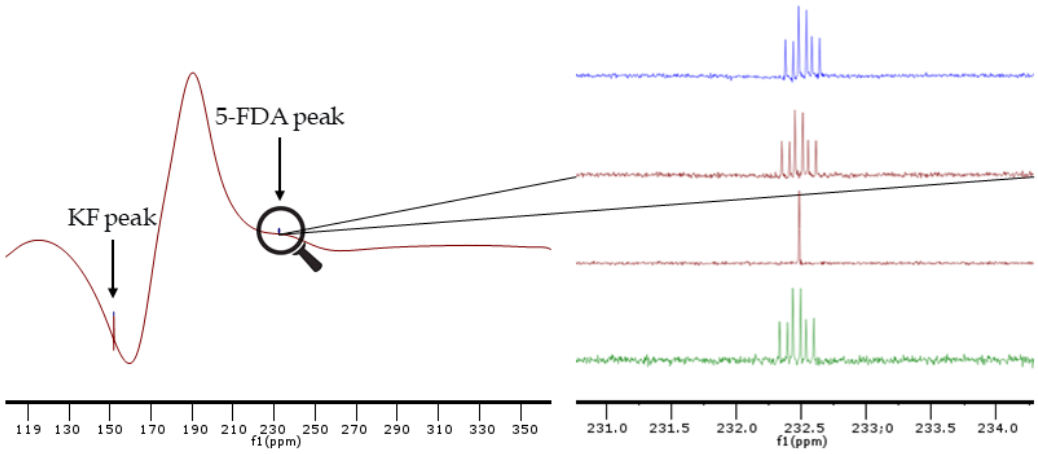


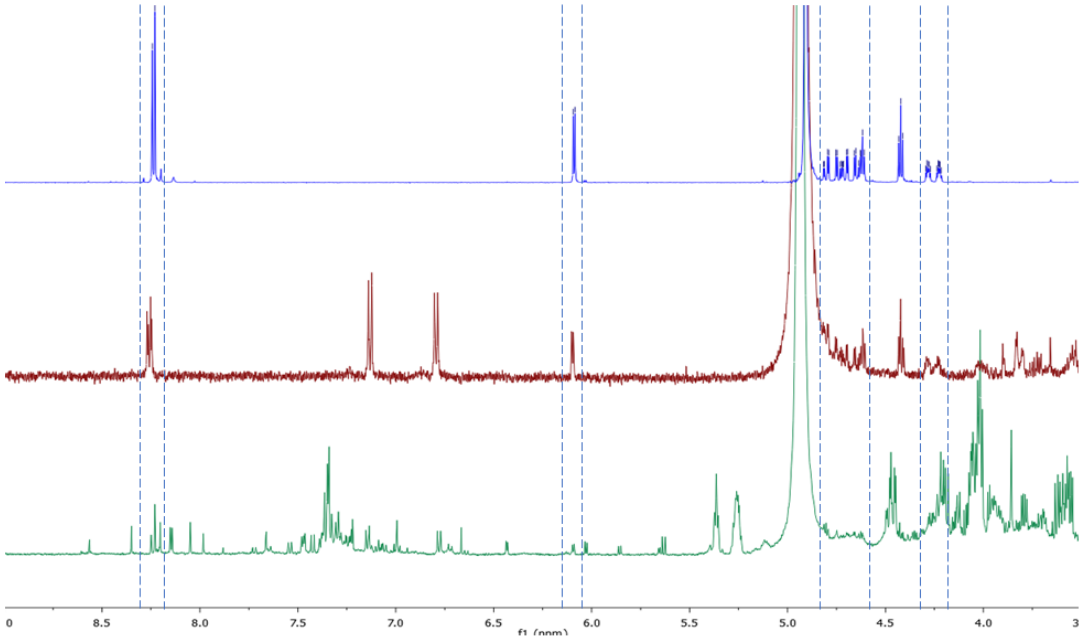
Figure 5.13: Left : Full ^{19}F-NMR spectrum of strain MK015. The potassium fluoride peak can be seen in -151ppm. Right: Zoomed ^{19}F-NMR spectrum around the fluorometabolite peak. 5'-FDA reference (blue), cell lysate of strain MK015 spiked with synthetic 5'-FDA (red, H-coupled and H-decoupled) and cell lysate of the same strain non-spiked (green).

in figure 5.13. The resulting coupled ^{19}F-NMR spectrum of the lysate clearly encompasses a multiplet which exhibits the same peaks with 5'-FDA. In order to further validate the identity of this compound, a reference sample of 5'-FDA and the lysate sample spiked with synthetic 5'-FDA were also tested. As shown in figure 5.13, the spiked sample showed no other peaks appearing in both coupled and decoupled spectra, but only an increase of the existing peaks of the lysate. This proves almost without a doubt the presence of 5'-FDA in the lysate. Analysis of strain MK019 shows similar results. Interestingly, no other multiplets appear in the spectrum suggesting the absence of other fluorometabolites in both MK015 and MK019. This result is somewhat surprising and shows either the incompatibility of 5'-FDA with both adenosine deaminase and PNP of *E. coli*, otherwise 5'-FDI or 5-FDRP would also be present, or that a compartmentalization of sorts exists in *E. coli*, resulting in the deaminase not

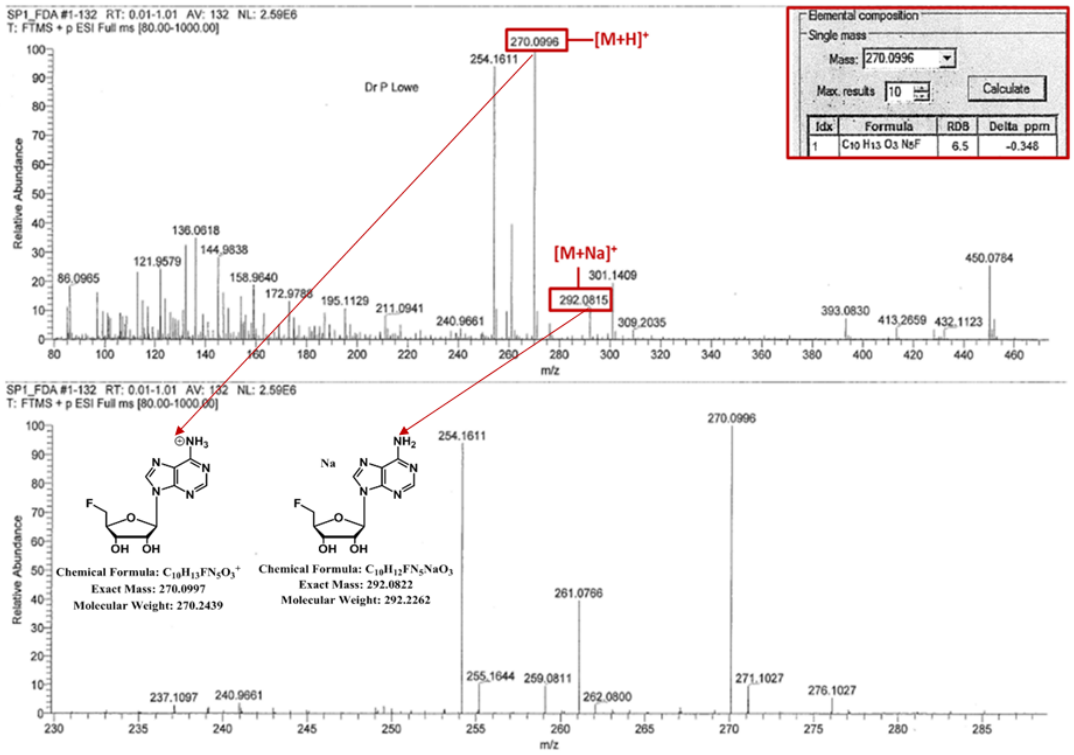
coming in contact with cytoplasmic 5'-FDA. The latter is also suggested for *S. cattleya* [43].

There are other possibilities such as the intermittent production of these fluorometabolites during incubation and subsequent degradation towards defluorinated products. However, absence of 5'-FDI in strain MK019 as well, diminishes this probability as 5'-FDI should at least be present since *E. coli* PNP, which processes 5'-FDI is deleted in this strain. All in all, 5'-FDA is produced and appears to persist inside the cells after a 24-hour incubation, regardless of the *deoD* gene being present or not. Although more experimentation is needed to confirm this result, fluorination does not appear to be hampered by the *E. coli* PNP *in vivo*, even though it is suggested that in a cell-free extract it does [52].

Although the above method is pretty much definitive of the identity of the compound, strain MK019 grown in 2 mM/50 μM of KF/SAM, which also demonstrated capacity for 5'-FDA production, was analysed with ^{1}H-NMR and HRMS in addition to ^{19}F-NMR. The results are shown in figure 5.14. The ^{1}H-NMR spectra (fig. 5.14a) consist of the crude lysate, the eluate from HPLC corresponding to the 5'-FDA peak (t_R = 14 min) and a synthetic 5'-FDA as reference. For both types of NMR analysis the solvent used was methanol-d4. The frequencies used for ^{1}H-NMR and for ^{19}F-NMR were 500.1 MHz and 470.6 MHz respectively. The multiplets appearing in the pure 5'-FDA sample are also clearly present in the HPLC eluate along with some unidentified peaks of a compound that is co-eluted in the same fraction. Some of the 5'-FDA peaks can also be discerned even in the crude lysate. For HRMS analysis, only the eluate was used and the highest peak appearing in the spectrum corresponds to the predicted m/z = 270 for cation $[M+H]^+$ with M equal to the mass of 5'-FDA. There is also presence of the peak with m/z = 292 for the sodium cation $[M+Na]^+$. It is therefore validated with 3 different spectral methods that the only fluorinated compound being produced intracellularly is 5'-FDA. As expected, the negative



(a) ^{1}H-NMR spectra of 5'-FDA synthetic sample, isolated HPLC fraction and crude lysate



(b) HRMS spectrum of the HPLC eluate

Figure 5.14: Top: 5'-FDA synthetic reference (Blue), isolated HPLC fraction (t_R=14 min, corresponding to known the retention time of 5'-FDA, Red) crude cell lysate of strain MK019 (2 mM KF) (Green). Bottom: HRMS analysis of isolated HPLC fraction.

control strains MK016 and MK020 grown in conditions 15 mM/250 μM KF/SAM did not reveal any peaks apart from KF in their ^{19}F-NMR spectra.

Due to the profound impact from establishing *in vivo* fluorination in *E. coli*, the possibility of production of 5'-FDA from released enzyme because of incomplete denaturing should be dismissed. For this test, the same conditions of growth were employed for strains MK015 and MK019 but with the absence of KF. The pellets were then normally lysed as described above but in the prewarmed aqueous ethanol, 15 mM of KF was added. If 5'-FDA detected in the previous experiments was produced due to incomplete denaturing then 5'-FDA would also be detected under these conditions. ^{19}F-NMR spectra did not reveal the presence of other fluorinated molecules apart from KF, concluding the series of experiments for the establishment of *in vivo* production of the fluorometabolite.

5.4.3 Contribution of modifications and quantitative analysis of 5'-FDA in the production strains

Following the confirmation of 5'-FDA production from the modified *E. coli*, in particular strains MK015 and MK019, it was evident that deletion of PNP does not affect intracellular 5'-FDA accumulation. At that point the question was raised which modifications and conditions contribute in the realization of *in vivo* fluorination. To this end, the different modifications were tested in an iterative manner. Initially, strains MK004, MK010 (2 mM KF/NO SAM) and MK015, MK019 (2 mM KF/500 μM SAM) were tested in otherwise same conditions as described in section 5.4.1. Strain MK004 was not expected to show production of 5'-FDA as in a concentration of 2 mM KF, fluoride anions would not be able to persist in the cell with a functioning CrcB channel. Surprisingly, no production was seen even in strain MK010, which means that intracellular SAM produced by the strain is either not enough for starting the reaction, or that it is concentrated

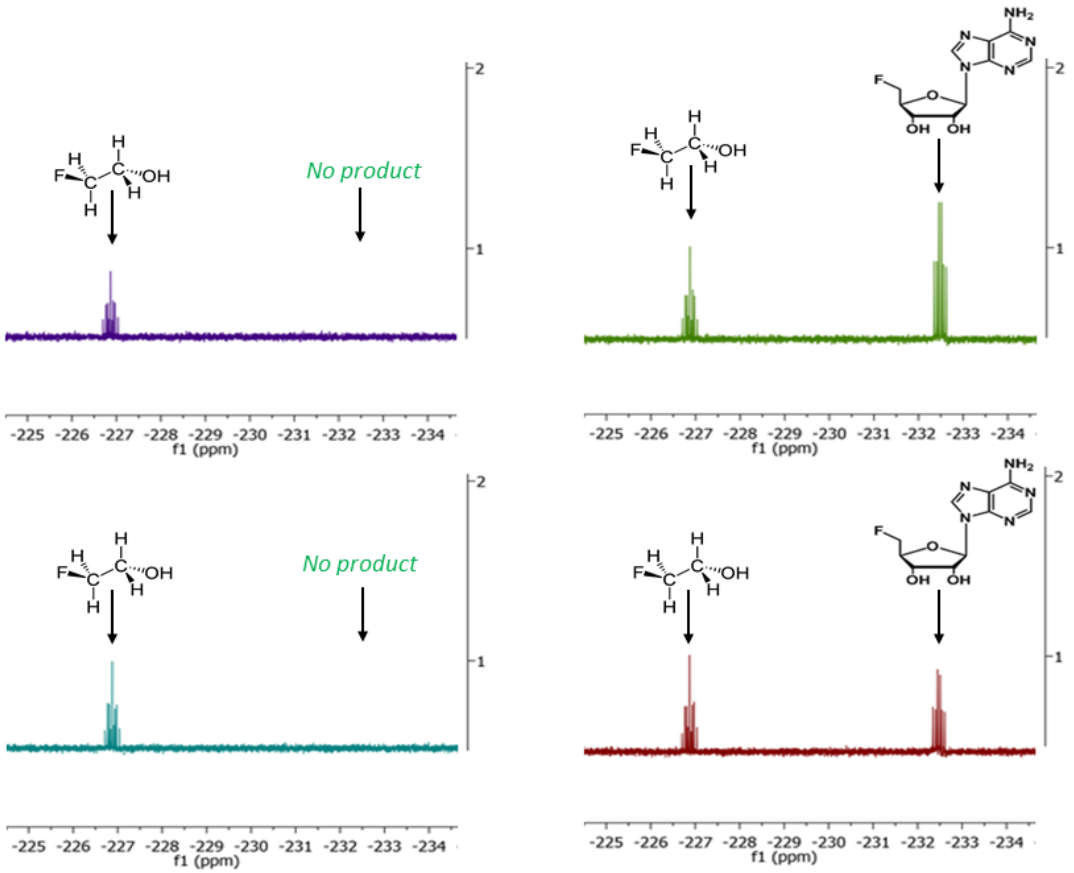


Figure 5.15: ^{19}F-NMR spectra of cell lysates to test for the production of 5'-FDA *in vivo* after 24h of incubation with 2 mM of KF. The modifications were tested in an iterative fasion. Lysates were spiked with 2-Fluoroethanol as a reference standard (-227 ppm). Upper-Left: Strain MK004 (No SAM), Lower-Left: Strain MK010 (No SAM). Upper-Right: Strain MK015 (500 µM SAM). Lower-Right: Strain MK019 (500 µM SAM). The multiplet appearing in -232.5 ppm belongs to 5'-FDA. The product cannot be detected on the strains lacking the SAM transporter.

in a particular sub-cellular location and out of reach for cytoplasmic fluorinase. To determine the intracellular concentration of 5'-FDA in the strains that showed 5'-FDA production, scaled-up quantities of lysates obtained with the procedure described in 5.4.1, were freeze-dried (liquid nitrogen) and subjected to ^{19}F-NMR as follows: First of all, preliminary ^{19}F-NMR experiments were performed using a solution of synthetic 5'-FDA and 2-fluoroethanol at the same concentration in

methanol-d4, in order to identify a sufficient relaxation delay. The longest delay between the fluorinated products was incorporated into the ^{19}F-NMR parameters to allow for accurate comparison of the fluorine peaks of 2-fluoroethanol and 5'-FDA. The freeze-dried lysates were added 700 µL of methanol-d4 spiked with a known concentration of 2-fluoroethanol. The suspension was vortexed and sonicated to ensure dissolution of 5'-FDA, then centrifuged for 5 min at 12,000 rpm to remove any precipitate. The solution was then analysed by ^{19}F-NMR on a Bruker Advance 500 instrument (NS = 220, D1 = 15, O1P = -229.75). The peaks for 2-fluoroethanol and 5'-FDA were integrated and the ratio was used to determine the 5'-FDA concentration. The latter was then used along with Dry Cell Weight (DCW) measurements to obtain mg/g values of 5'-FDA production.

The results from one replicate for strains MK004, MK010, MK015 and MK019 are shown in figure 5.15. The intracellular concentration of 5'-FDA that has been achieved, as determined from 2 replicates, is 0.73 (± 0.33) mg/g DCW in strain MK015 and 0.59 (± 0.03) mg/g DCW in MK019. The incubation conditions were 2 mM KF, 500 µM SAM, 24 h in 37 ^{o}C. Assuming a 10^{-15} L of volume per cell, 10^{9} cells/mL in 1 OD and 0.33 g/L of DCW in the same OD, yields 0.33mg/10^{9} cells, therefore 33x10^{-14}g of DCW/cell and subsequently 24.1x10^{-17}g of 5'-FDA per cell. In terms of molarity, this is equivalent to 24.1x10^{-17}g/10^{-15} L, hence 241 mg/L. 5'-FDA molar mass is ≈ 270 g/mol, therefore the intracellular concentration is 0.89 (± 0.4) mM in strain MK015 and similarly, 0.72 (± 0.04) mM for strain MK019.

5.4.4 Identification of the critical modifications for *in vivo* fluorination

The total absence of 5'-FDA in strains lacking the SAM transporter, warranted more exhaustive testing with combinations of strains/conditions in order to

unravel or dismiss secondary effects. Three concentrations of SAM were tested in strains MK015 and MK019 (20 μM 100 μM and 500 μM). In addition, strain MK008 was also tested with 500 μM of SAM and strains MK004, MK010 were tested in 0 and 500 μM of SAM. The results from these experiments are presented in a manner were production or non-production is determined (table 5.1). From the results, it is evident that both the deletion of the *crcB* gene and SAM transporter expression are key modifications for enabling *in vivo* fluorination in an *E. coli* host expressing a fluorinase enzyme. This synergy of actions is profound by the non-productivity of 5'-FDA when each of these modifications exist separately in a strain combined with fluorinase expression (MK008 and MK010). Strains with combinations including the PNP deletion ($\Delta deoD$) were not tested as its absence did not seem to exert a positive effect on the production of 5'-FDA. Testing of a saturating concentration of SAM (500 μM) in strains lacking the SAM transporter also dismissed the possibility of a secondary effect other than the influx of SAM from the transporter (e.g. higher fluoride accumulation from alteration of pH). Additionally, 20 mL of S/N collected from 5'-FDA producing strains MK015 and MK019 in 2 mM/500 μM KF/SAM, and non-producing strains was freeze-dried and subjected to ^{19}F-NMR. No 5'-FDA (or any other fluorometabolite) was observed, indicating that the product accumulates in the cytoplasm and is not exported.

Table 5.1: Validation of *in vivo* fluorination with 2 mM of KF in different strains and SAM concentrations. Lysates collected from all these incubations where tested in 3 replicate experiments. "YES" or "NO" cells have been tested with ^{19}F-NMR and showed presence or absence of 5'-FDA respectively. Cells denoted as "N/A" were not tested.
(*) *Strains MK016 and MK020 contain the negative control pET-151(-) and they were tested only once in conditions 15 mM/250 μM KF/SAM.*

		SAM concentration											
		NO SAM			20 μM			100 μM			500 μM		
Strain	MK004	NO	NO	NO	N/A			N/A			NO	NO	NO
	MK008	N/A			N/A			N/A			NO	NO	NO
	MK010	NO	NO	NO	N/A			N/A			NO	NO	NO
	MK015	N/A			NO	NO	NO	YES	YES	NO	YES	YES	YES
	MK016	NO*											
	MK019	N/A			YES	YES	NO	YES	YES	NO	YES	YES	YES
	MK020	NO*											

Chapter 6

Raman Spectroscopy of Fluorometabolites

RS is gaining increasing attention as a non-invasive, non-destructive, label-free method for the determination of chemical composition in the context of live cells or tissues. A small part of the present work is therefore dedicated in exploring the potential for characterizing and possibly identifying the substrates and products of enzymatic fluorination with this spectroscopic method based on fingerprint spectra for these compounds.

This chapter presents previously unreported Raman spectra for SAM chloride dihydrochloride, 5'-FDA and 5'-ClDA. Novel findings include:

1) The ability to monitor the spontaneous degradation of SAM with Raman.

2) Raman enabled determination of SAM uptake from *E. coli* cells.

6.1 Experimental set-up

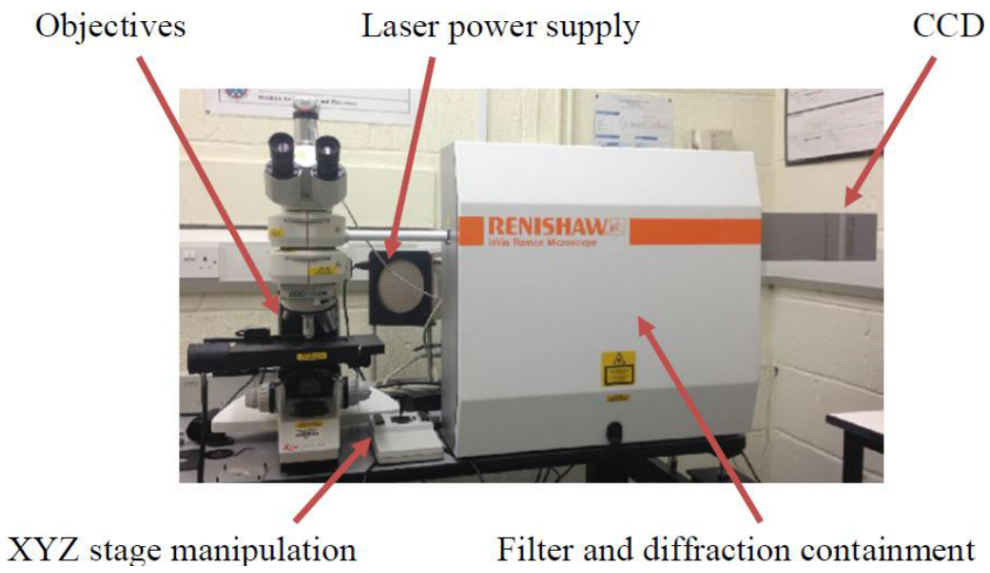
This section presents the experimental set-up for Raman spectral acquisitions, spectral and spatial resolution measurements, as well as the methods utilized for cosmic ray spike and baseline removal.

6.1.1 Instrument and lens

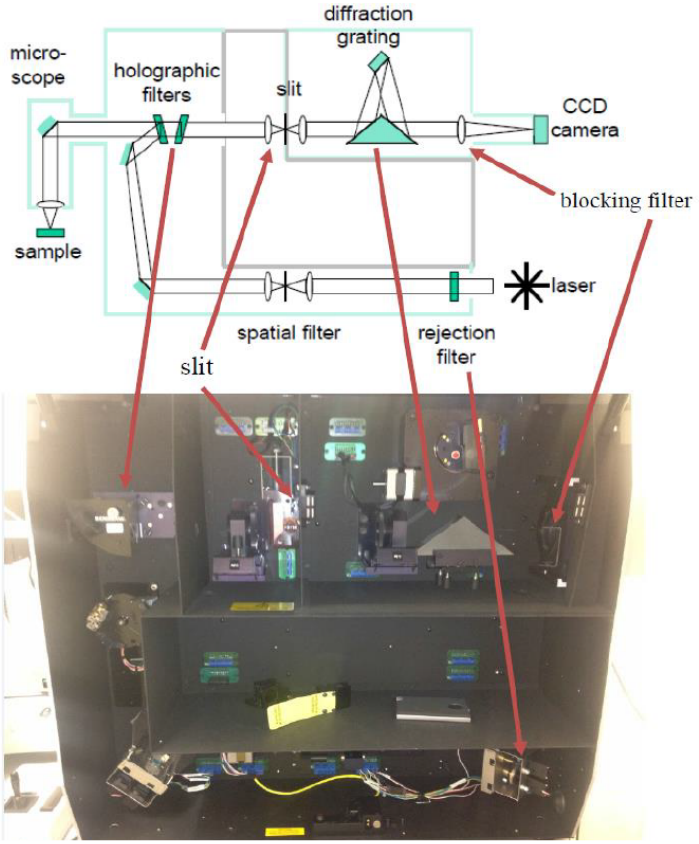
The instrument used for the spectral acquisitions is a Renishaw InVia^{TM} Raman Microscope specifically designed for the acquisition of Raman spectral signatures from biological samples (see figure 6.1). The excitation light that this instrument utilizes has a 785 nm wavelength (near infra-red range) and an output power of 60 mW. The power of light that reaches the sample however, was measured to be lower, around 40 mW due to the cumulative effect of small aberrations of the various elements along the optical path. These low-intensity and near infra-red wavelength photons do not ionize the sample and are therefore appropriate for non-destructive spectral acquisition. The only concern is thermal damage potentially caused to living cells from long exposure times. In the following series of experiments, this is not a concern as measurements are performed with limited exposure times (10 - 30 seconds) and the majority of spectral acquisitions involve inert samples. The lens used was a Leica HC PL FLUOTAR 50x/0.80 DB. The moving stage features a minimum step of 0.5 μm in the XY axes (lateral) and 1 μm in the Z axis (axial).

6.1.2 Spatial Resolution

In confocal microscopy, the focused light beam spot size is representative of the spatial resolution that can be achieved. A small focal spot will generally yield a



(a) Renishaw Raman InVia^{TM}



(b) Layout and components

Figure 6.1: The instrument and layout used for the acquisition of Raman spectra. Taken from [10].

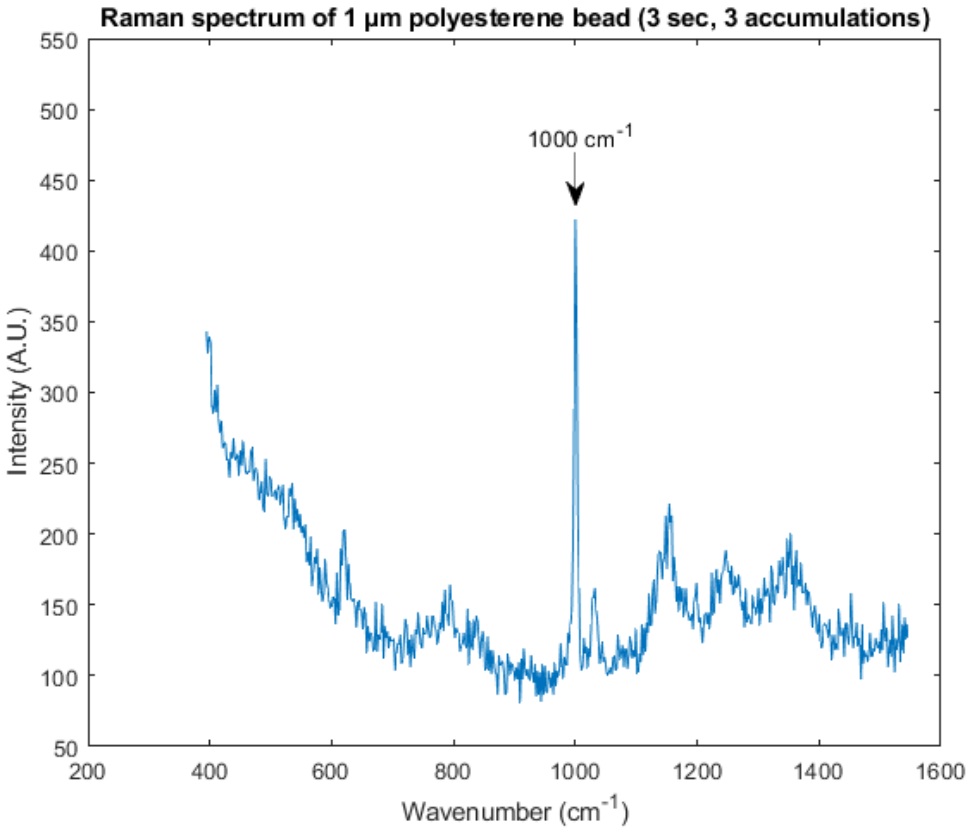


Figure 6.2: Spectrum of a polyesterene bead. The higher intensity peak appears in 1000 cm^{-1}.

better spatial resolution in the sample. There is a trade-off however in the signal intensity, as a smaller focal spot will encapsulate a proportionally lower volume from the sample. On the other hand, a larger spot size will encompass a higher volume and yield a higher signal, but will be harder to resolve differences in the sample. Many factors must be considered in choosing the optimal objective lens for each experiment such as the heterogeneity (as is the case for mammalian cells), and density of the sample. In this work, measurements involve homogeneous samples of pure powders, solutions and populations of dried or live bacteria, thus the measured spatial resolution is satisfactory. The theoretical minimum of axial and lateral resolution for an objective lens can be derived from the Rayleigh criterion (eq. 6.1).

$$R_{Lateral} = \frac{0.61\lambda}{NA} \qquad R_{Axial} = \frac{2n\lambda}{(NA)^2} \tag{6.1}$$

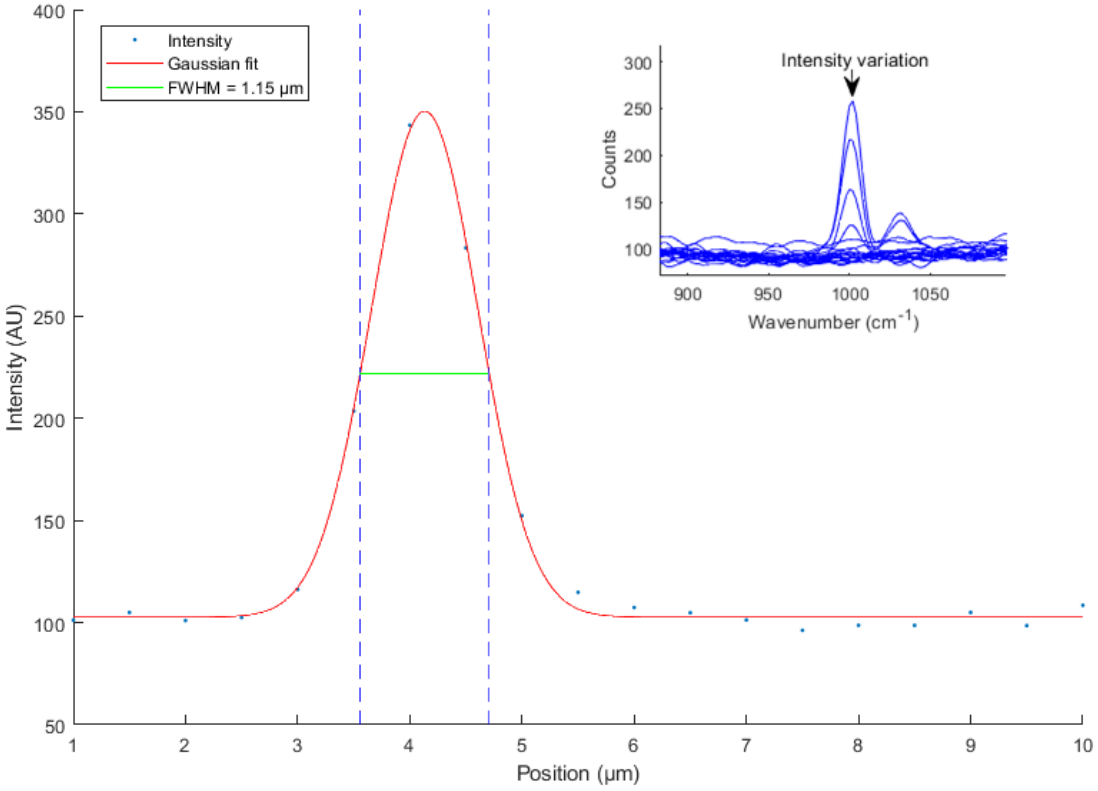
Substituting with $\lambda = 785$ nm and $NA = 0.8$ as per the instrument and lens specifications, the resulting theoretical resolution is:

$$R_{Lateral} = 0.6\ \mu\text{m} \qquad R_{Axial} = 2.45\ \mu\text{m} \tag{6.2}$$

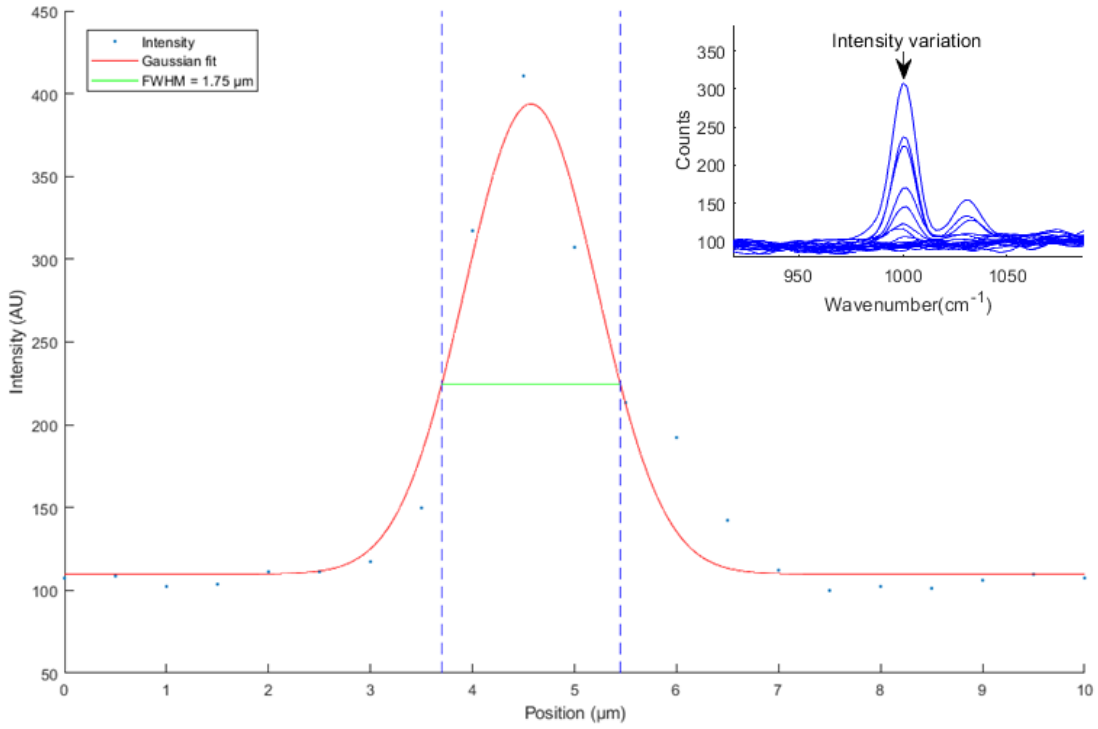
Where n is the diffraction index of air and equals 1. This is the diffraction-limited maximum resolution that can be achieved and is a quantum phenomenon. The real resolution is always lower (higher in terms of length) than this and is dependent on the quality of optical elements. Another aspect of measurement that could pose a limit for the spatial resolution are the minimum steps of the moving stage for X,Y and Z axis, which are 0.5, 0.5 and 1 μm respectively. In this case it is lower than the diffraction limit for this lens and does not constitute a limiting factor.

Experimental spatial resolution

For experimental determination of the spatial resolution, 30 μl of liquid containing small polyesterene spherical beads (1 μm) was poured onto a quartz substrate, subjected to a short centrifuge pulse for the liquid to be effectively dispersed and left to dry. The substrate was then placed onto the instrument stage and the laser was focused on an isolated bead with a sufficiently empty surrounding space. A spectral acquisition was obtained with a 2 second exposure and 10 accumulations to determine the best focus based on signal strength. The resulting spectrum is shown in figure 6.2. The highest intensity peak appears at ≈1000 cm^{-1}. A linear



(a) X-axis resolution



(b) Y-axis resolution

Figure 6.3: Determination of spatial resolution based on measurements of intensity with a moving focal spot.

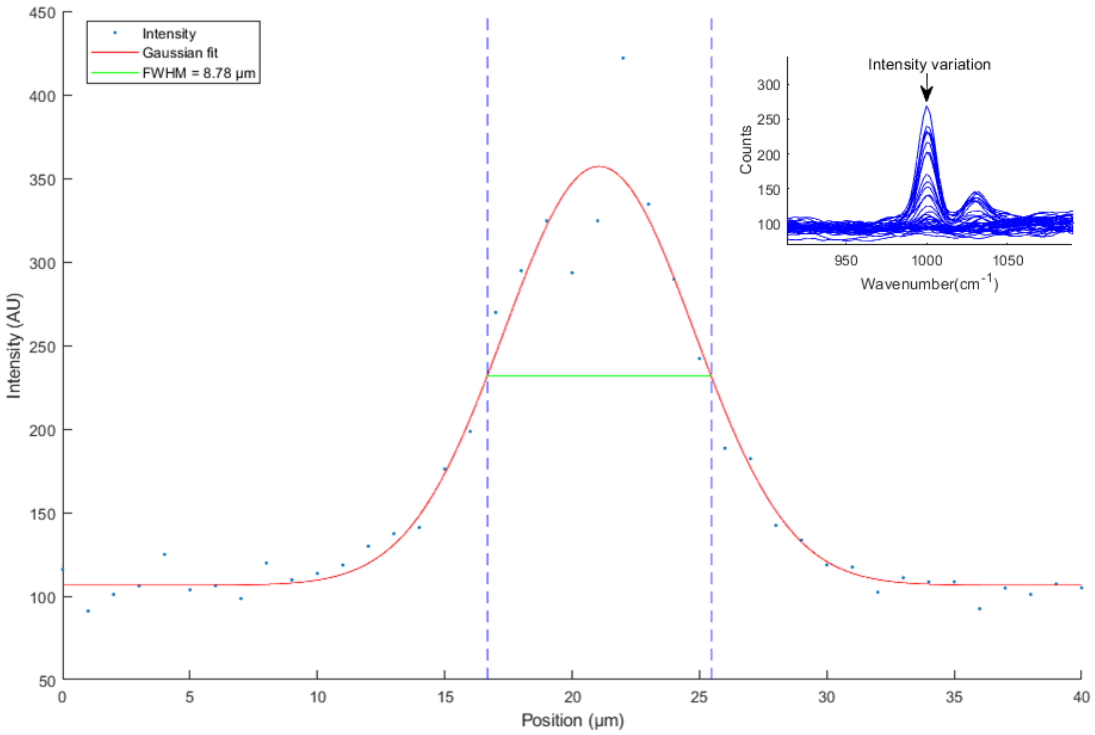


Figure 6.4: Z-axis resolution.

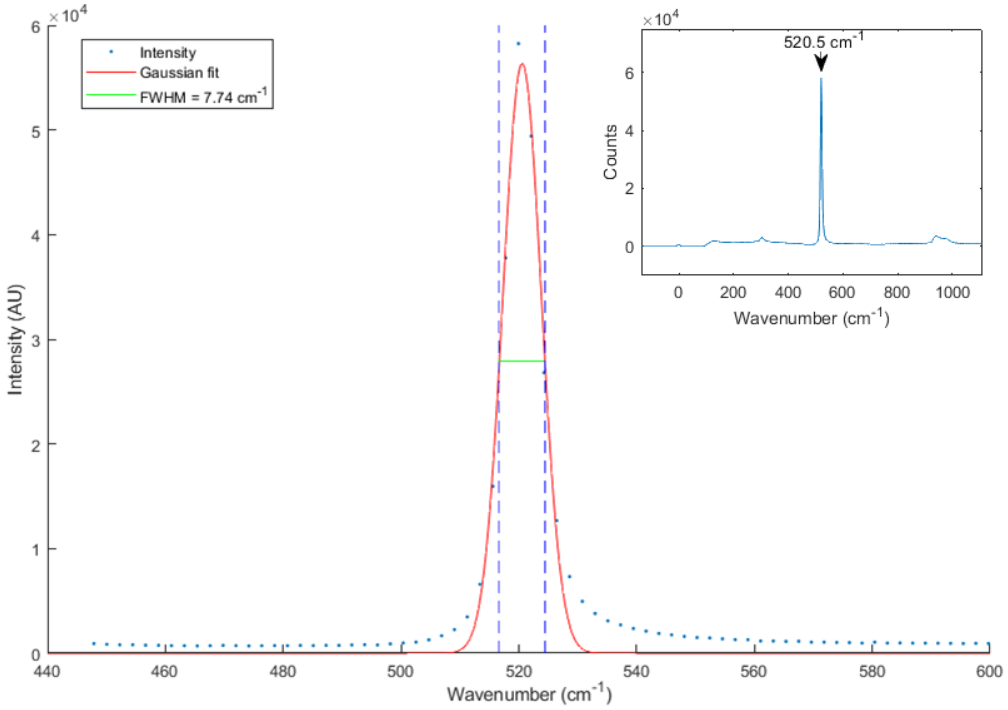
mapping measurement was set up overlapping the sphere by at least 10 times its diameter on each side. With every moving step, as the focal spot passes through the polyesterene sphere, the intensity increased reaching a maximum when the sphere was located at the centre of the focal spot and subsequently decreased as the spot moved away. The varying intensity of the highest peak (≈1000 cm^{-1}) from the movement of the focal spot along each axis are shown in figures 6.3 and 6.4. The Full-Width Half-Maximum (FWHM) resolution is calculated based on a Gaussian curve fit for each series of measurements. The X and Y axis feature a FWHM of 1.15 and 1.75 μm respectively. The highest FWHM length is considered as the lateral resolution and it is ≈3 times the minimum resolution length (eq. 6.2). The measured axial FWHM resolution is 8.78 μm and ≈3.5 times higher than the theoretical minimum for this lens.

6.1.3 Spectral resolution

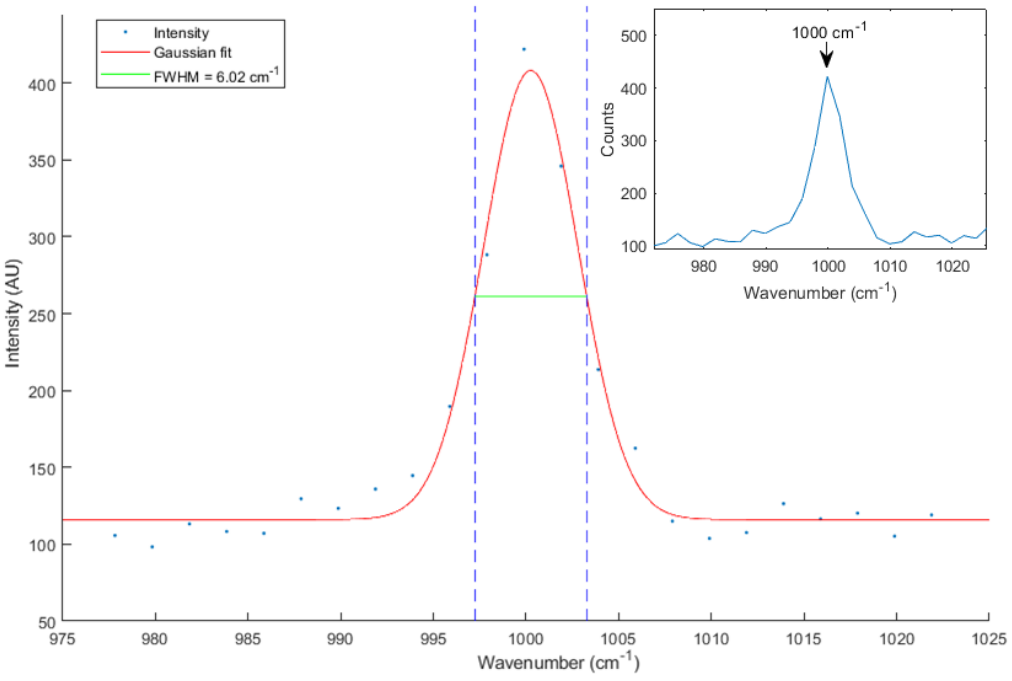
The spectral resolution of a Raman spectrometer is defined as the minimum wavelength difference at which two adjacent spectral lines (peaks in the spectrum) can be distinguished. There are several factors that determine the spectral resolution [154]:

- The natural spectral line widths of the substance itself. There is a lower limit which is tied to the uncertainty principle. Other factors also contribute to the natural line width such as the temperature, homogeneity and crystallinity of the sample.
- The number of grooves (lines) per millimetre of the grating. A higher number of lines will disperse the different wavelengths of the signal over a wider area in the CCD detector. One trade-off is that the spectral window will be lower for one acquisition. Thus, choice of grating must be accompanied with considerations specific for each sample.
- The detector slit size and the pixel size of the CCD.
- System specific diffraction effects, aberrations and the laser line width.

The spectral resolution was determined experimentally using FWHM by Gaussian curve fitting plus a non-zero constant. The latter was calculated in two substances, first using the peak that appears in ≈ 520.5 cm^{-1} from a silicon sample and second, using the peak in ≈ 1000 cm^{-1} from a polyesterene bead. The silicon sample is also used for the internal calibration procedures of the Raman instrument. The grating utilized for these and all subsequent measurements contained 600 l/mm (Figure 6.5). There is a difference in the measured FWHM resolution between substances, in particular, 7.75 cm^{-1} in silicon and 6.02 cm^{-1} in polyesterene. This can be justified by the fact that there is a difference in wavenumbers (520.5 cm^{-1} vs 1000 cm^{-1}) an effect also illustrated in [155].

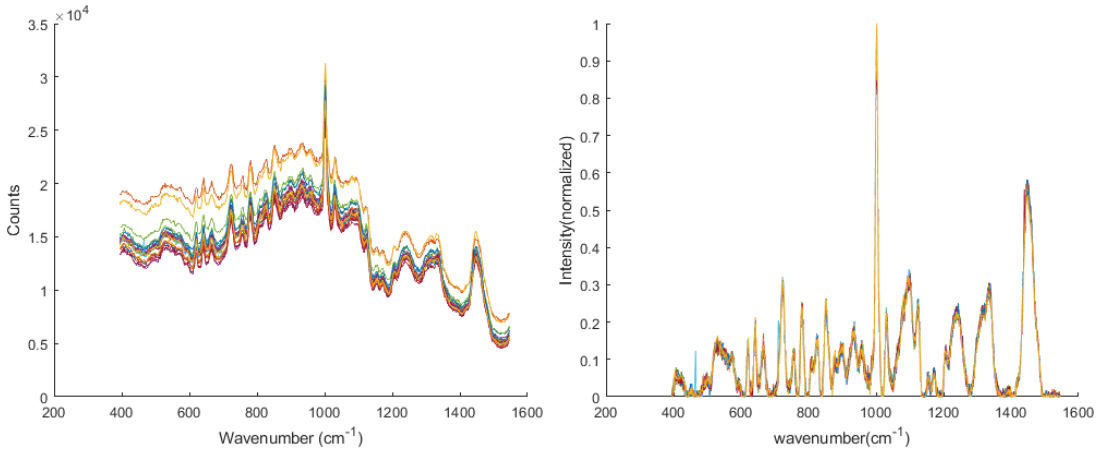


(a) FWHM resolution of silicon peak (520.5 cm^{-1})



(b) FWHM resolution of polyesterene peak (1000 cm^{-1})

Figure 6.5: Determination of spectral resolution from intensity peak variation within a single acquisition.



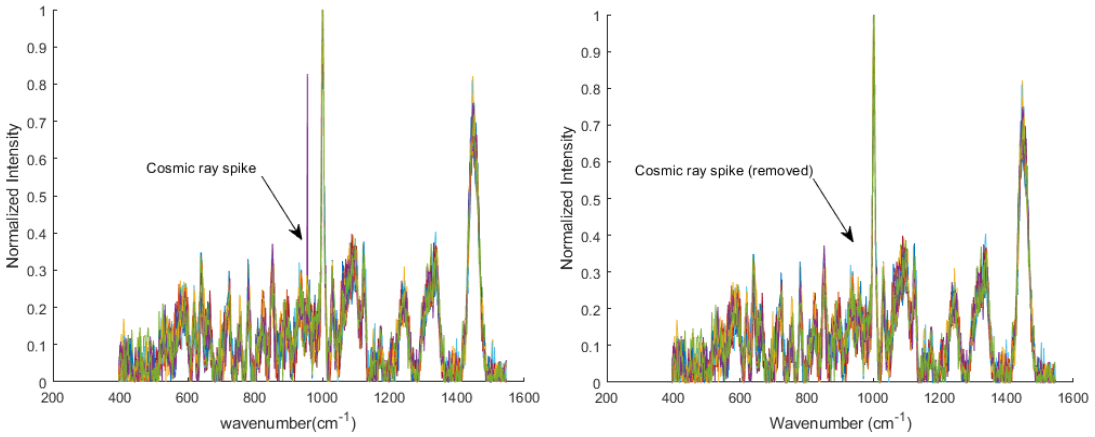
(a) Raw spectra

(b) Spectra with the baseline removed and normalized to the highest intensity peak ($\approx$ 1002 cm^{-1})

Figure 6.6: Example application of the baseline removal algorithm.

6.1.4 Baseline removal

Raman spectra, apart from the peaks that are produced from scattered, Raman-shifted photons that are particular to vibrational modes of molecular bonds being excited, contain a smooth signal which appears due to the absorption of photons by atoms in the sample and excitation of their electrons in higher energy states. The atoms subsequently fall into the ground state and re-emit photons of a particular frequency. This is a different effect from Raman scattering, termed autofluorescence and manifests itself as a baseline in the spectrum. It is more pronounced with higher energy photons comprising blue or green wavelength laser beams. With a 785 nm laser used in this study the effect is not detrimental to the signal/noise ratio, but still needs removing as each acquisition suffers from different intensities of autofluorescence (see figure 6.6a). Usually, preprocessing steps include algorithms for removing this baseline from the acquisition. In this case, an automated small window moving average approach is utilized and the algorithm was kindly provided by the author of the original work [156]. This



(a) Spiked spectra **(b)** Despiked spectra

Figure 6.7: Example application of the spiking removal algorithm.

algorithm provides a highly robust approach for effectively removing baselines from spectra in which even manual removal would be problematic. The algorithm works in an iterative manner by performing peak-stripping/window-size enlarging and then determines the optimal number of iterations based on the areas below the signal line being removed. It is important to note that the spectra acquired for section 6.2, required an adaptation of the algorithm. A first pass of the algorithm was used to extract the number of iterations for each individual spectrum and the average of this number was used statically for the second pass. This enabled the correct alignment of spectra (shown in figure 6.6b).

6.1.5 Cosmic ray spike removal

The detector used during acquisitions of Raman spectra is a CCD camera which consists of an array of pixels each with a size of 22 μm x 22 μm. The detector collects Raman-shift generated photons from the light beam re-emitted from the sample, effectively separated from the grating in its constituent wavelengths and dispersed to different pixels. Cosmic rays are highly energetic nuclei mostly

originating from outside the solar system that continuously bombard the earth. During long exposure/acquisition times, there is a probability that a cosmic ray will "hit" a particular pixel in the CCD and this event is recorded by the detector, appearing as a cosmic ray spike in the final spectrum. There are several pre-processing methods for removing these undesired artefacts. The technique used here is an adaptation from the method laid out in [157] and requires a number of replicated "similar spectra". Specifically, the measurement raw file is an array of wavenumbers and their corresponding intensities. With a number of spectra at hand, one of them will occasionally exhibit a cosmic ray spike at a particular wavenumber that others will not. This spike will constitute an outlier and can be removed by setting an upper-limit of 4 times the standard deviation of the intensities in this particular wavenumber. Measurements that do not satisfy this limit will be recognized as spikes and levelled off. The algorithm should ideally be applied in spectra with removed baseline. An application example is shown in figure 6.7.

6.2 SAM characterization

The spectra and assignments of bonds for SAM will be presented here. SAM degradation will be illustrated with measurements taken in different days and the decrease of a particular peak will be linked to this effect. The uptake of SAM at high molarities from the function of the SAM transporter will be argued for based on the statistically significant difference of the highest peak present in SAM, also appearing in the spectral signatures of strains with SAM transporter expression grown in the presence of SAM.

The importance of SAM in biological processes has been laid out in section 4.4. Surprisingly, almost two decades after the burst of RS application in a biological context, a Raman spectrum of this important molecule has yet to be reported

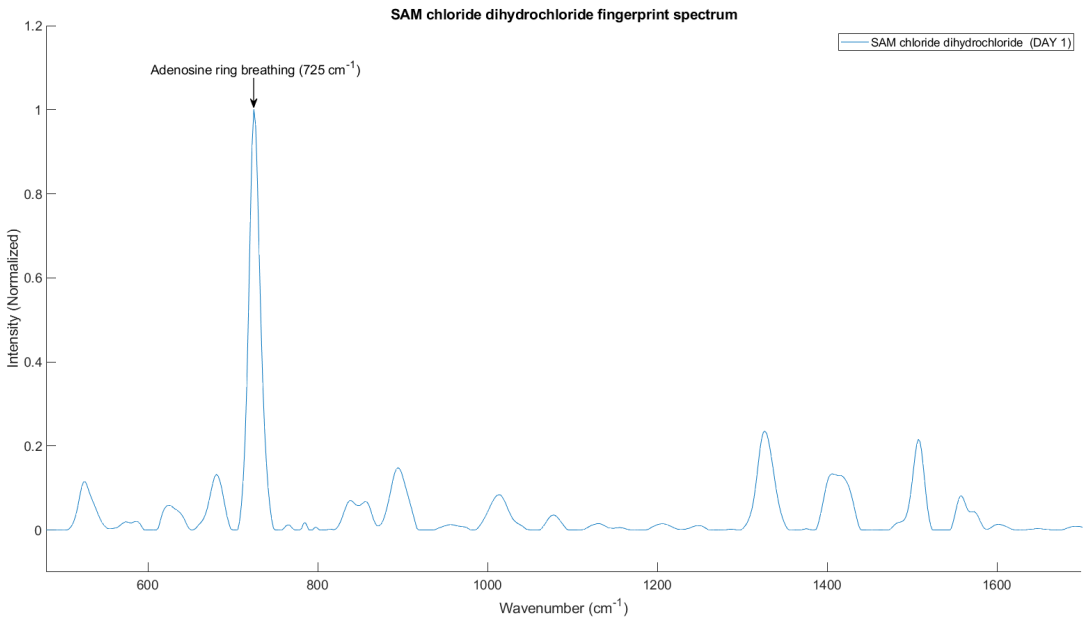


Figure 6.8: Mean spectrum of 48 spectral acquisitions of SAM chloride dihydrochloride powder in fingerprint region baseline removed and normalized (10 second exposure).

(to the best of the author's knowledge). The reason behind this profound gap could be the fact that SAM is a cation and unstable outside of the intracellular environment, therefore measurements of this sort are inherently problematic. In market, it can be obtained as a salt. For all SAM-dependent experiments performed as a part of this project, the molecule in this chemical form (obtained from SIGMA) was used. The molecule, albeit being more stable as a salt, it stills suffers from instability which is temperature-dependent. The mechanism of spontaneous cleavage is mentioned in [152]. SIGMA guarantees 75% of purity at the time of purchase and reports that it can lose up to 10% of purity per day in room temperature [158].

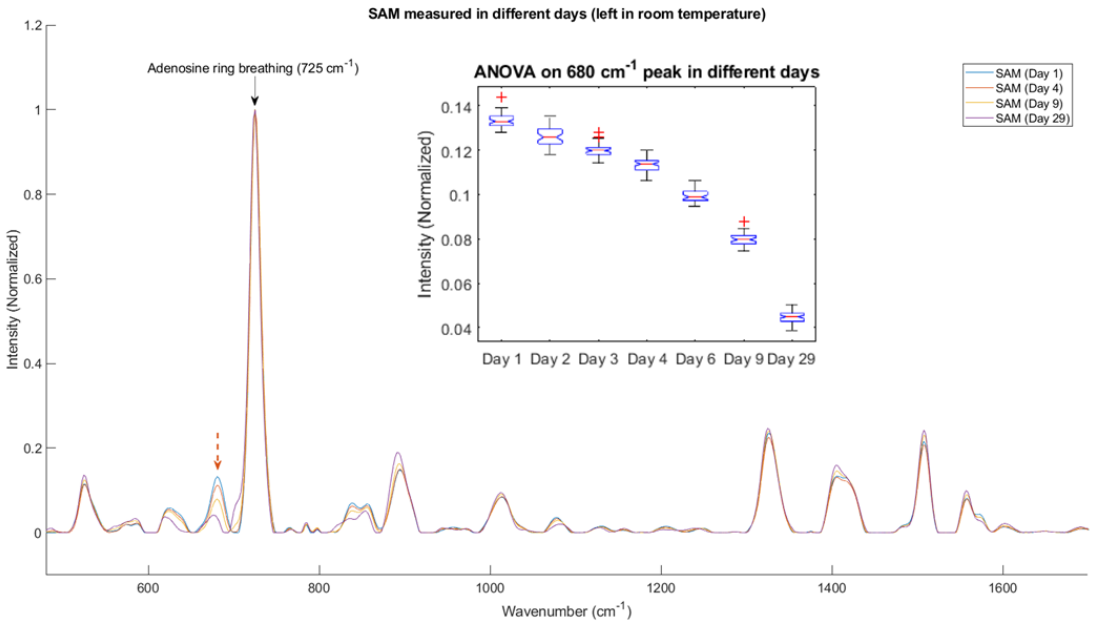


Figure 6.9: Spectral acquisitions of SAM in different days after exposure in 25 ^{o}C. ANOVA shows statistical significant differences between all days and a decreasing trend.

6.2.1 Impure SAM spectrum and degradation monitoring

A small quantity of the product, which appears as an off-white powder, was placed on a quartz substrate and a number of Raman measurements were performed with the above setup, in different locations of the substrate with powder presence (10 second exposure - 3 accumulations). The spectra were subjected to anti-spiking, baseline removal and normalization to the highest peak which appears at approximately 725 cm^{-1}. An average spectrum is shown in figure 6.8. Error curves are too narrow to be visually distinct and are not included.

It is worth noting that because of SAM instability a large percentage (up to 25%) of the substance was expected to have transformed to the degradation products (mostly 5-methylthioadenosine and homoserine lactone) at the time of the first measurement. Because of this instability, most peaks cannot be assigned with certainty to SAM-specific bonds. The highest peak that appears in 725

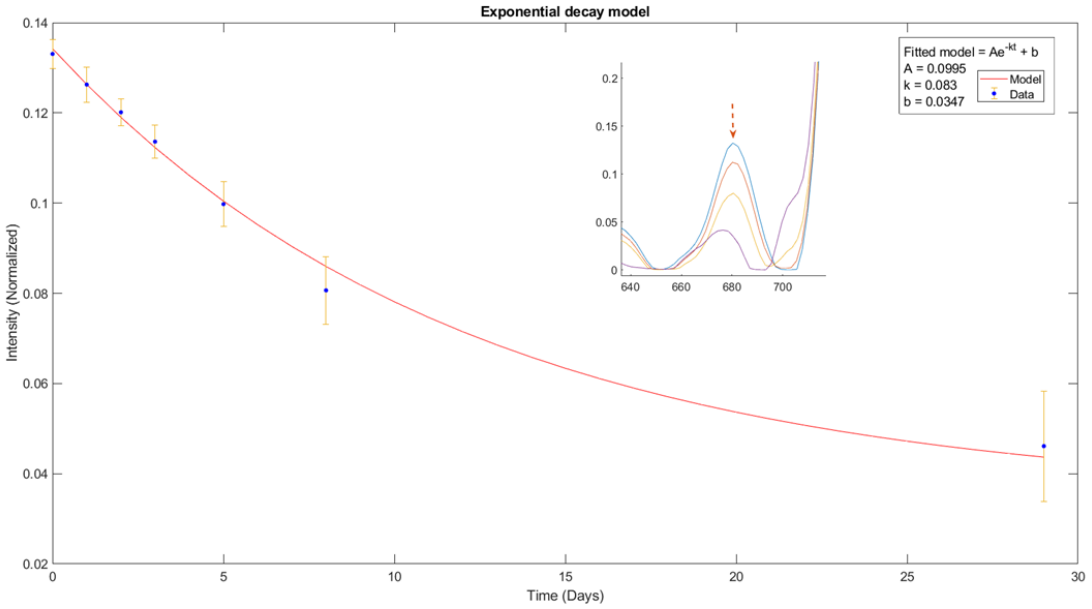


Figure 6.10: Exponential decay model based on the decrease in intensity of the peak in 680 cm^{-1}.

cm^{-1}, however, can be assigned to the adenosine ring breathing mode (see also [159] and section 6.3). The instability of SAM fuelled the following experiment: The SAM powder residue used for the first spectral acquisition was left in a closed but aerated box in room temperature and measurements were taken in different days. Specifically, 7 series of acquisitions (48 measurements for each) were acquired in days 1, 2, 3, 4, 6, 9 and 29 including the one presented above. Several peaks were monitored for changes in their relative intensity to the 725 cm^{-1} peak (normalization essentially yields ratios w.r.t. this peak). This peak constitutes a solid choice for normalization not only because it is the most intense but also the adenosine ring breathing mode producing this peak persists in the degradation products. Indeed, several peaks undergo changes in their intensities, others with an increasing trend and others show a decrease. There are also slight shifts in the wavelengths of some peaks. The peak that appears in 680 cm^{-1} undergoes a consistently significant (ANOVA) decrease between all measurements. By studying the degradation products and in consistency with literature [160],

this peak is assigned to the C-S bond connecting the adenosine part with the L-methionine part of SAM (figure 6.10). In effect, the degradation products contain 2 C-S bonds out of 3 in the initial molecule.

An exponential decay model was subsequently fitted using the intensity ratio and accounting for the variance of each measurement. The result is shown in figure 6.10. It is interesting to quantify the half-life of SAM in room temperature and $t_{9/10}$ (time were 90% of SAM remains intact). Results are as follows:

$$t_{\frac{1}{2}} = \frac{ln(2)}{k} = 8.35\ d \qquad\qquad t_{\frac{9}{10}} = \frac{ln(\frac{10}{9})}{k} = 1.27\ d \tag{6.3}$$

The result for $t_{9/10}$ shows 10% of SAM is degraded in 1.27 days, therefore it is in agreement with reported purity loss from SIGMA, serving as a validation for the utility of this approach.

6.2.2 Raman enabled detection of SAM uptake in *E. coli* cells

In this section, the detection of SAM uptake from *E. coli* grown in the presence of SAM will be investigated, based on the measurement of characteristic Raman peaks, which are prevalent in the spectrum of SAM chloride dihydrochloride powder (see previous section), also found in spectra acquired from *E. coli* featuring SAM transporter expression. The line of thought is adapted from [161] where the detection of xylitol uptake in *E. coli* cells is described but instead of the pictorial comparison that the aforementioned study employs, a statistical significance (ANOVA) approach will be followed here with a focus on the largest intensity peak of SAM ($\approx 725\ cm^{-1}$). Two comparative experiments were designed for the

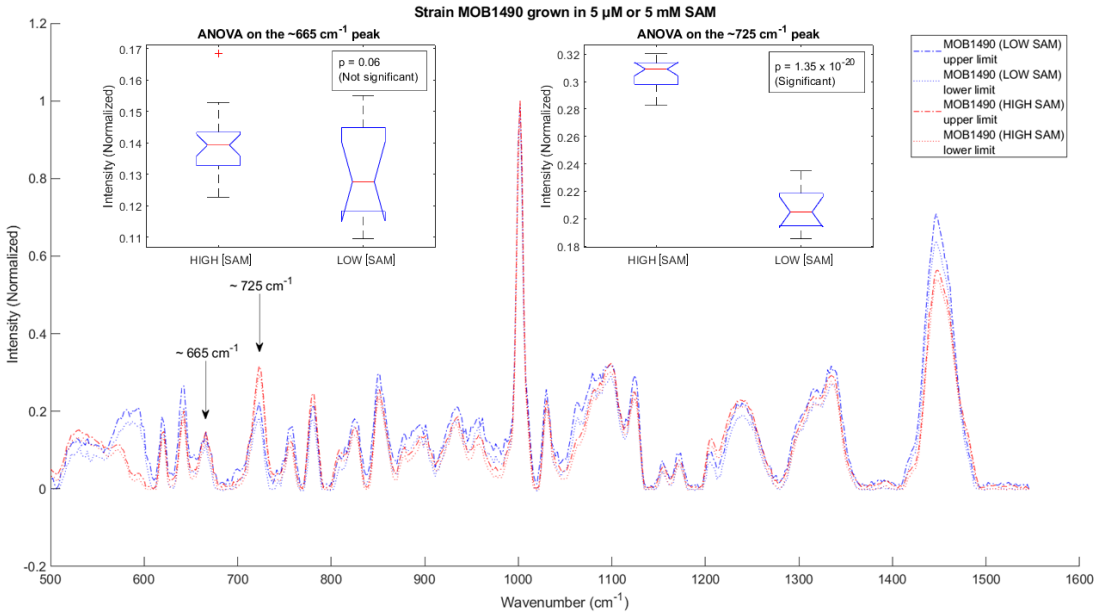


Figure 6.11: Spectral ranges of dried out cells from strain MOB1490 grown in 5 µM or 5 mM SAM. The highest SAM peak shows statistically significant higher intensity in the high SAM concentration.

identification of this effect. In the first experiment, 2 liquid (LB) cultures (3 ml in 15 ml falcon tubes) of strain MOB1490 were grown containing 1000-fold different SAM concentrations, limiting SAM (5 µM, non-optimal growth) and excessively high SAM (5 mM). The cultures were left to incubate for 24 hours and then centrifuged. The S/N was removed carefully and the pellet was re-suspended in an equal amount of MiliQ water, and then centrifuged again. This washing round was performed one time in the low SAM and four times in the high SAM concentration to ensure that no residual SAM from the growth medium was left in the samples. Finally, 5 µl of wet pellets were poured onto CaF_2 substrates and left to dry. Raman acquisitions were then taken from the sample in randomly chosen XY positions of the substrate where dried cell residue was located. The resulting spectra are shown in figure 6.11.

The peak of interest in the acquired spectra corresponds to the highest peak

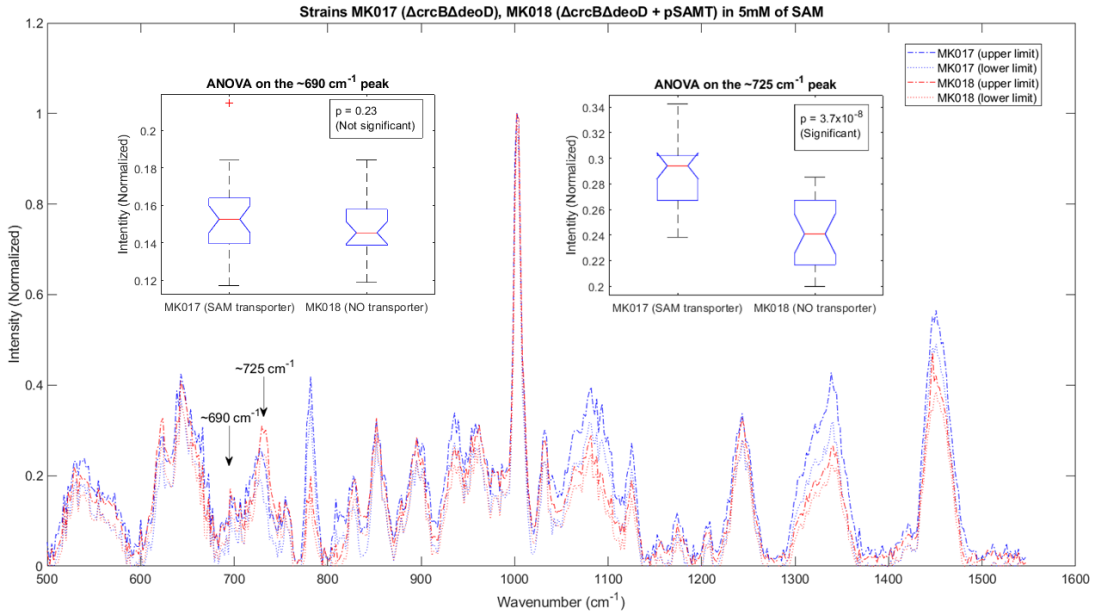


Figure 6.12: Spectral ranges of live strains MK017 and MK018 grown in 5 mM SAM and then deposited in agar. The highest SAM peak shows statistically significant higher intensity in the MK018 strain.

appearing in SAM ($\approx$ 725 cm^{-1}) which corresponds to adenosine ring breathing mode. In the spectra corresponding to growth in higher SAM concentration, the peak is clearly more intense with indisputable statistical significance shown by ANOVA. The peak which appears closer to 680 cm^{-1} (the signature peak of SAM), though exhibiting a higher average intensity, its mean compared to that of low SAM is not significantly different. This result provides strong evidence that the difference in the 725 cm^{-1} peak is indicative of more SAM being absorbed from the SAM transporter but the difference cannot be attributed entirely to the molecule of SAM as the degradation product 5-MTA will also exhibit this peak and there are transporters that could transport this molecule in *E. coli*. It was therefore deemed necessary to conduct another comparative experiment, this time the difference being the expression of SAM transporter. To this end, LB liquid cultures of strains MK017 and MK018 were grown as described above but they were both added 5 mM of SAM and left to incubate for 2 days. This would allow

any contribution from the degradation products to be pronounced. The pellets were collected and washed 4 times as explained above and then were put to agar plates and left to dry. Raman spectra were then acquired as shown in figure 6.12. The peak assigned to adenosine ring breathing still shows a signifficant but lower difference in intensities between strains. Combined with the previous result it can now be stated that the difference in intensity of this peak is indicative of SAM uptaken by the bacteria. The mean value of the peak closer to ≈680 cm^{-1} is higher but still does not show statistical significance according to ANOVA. The latter is representative of the limits of this particular spectroscopic method.

6.3 Raman spectra of the fluorinase products

The above results show the potential of RS to identify interesting intracellular compounds accumulating in bacterial cells. Therefore, RS as a selection method for improved variants of a particular enzyme can be envisaged. It has also been shown that simply utilizing Raman acquisitions suffers from the limit of detection of this particular method which is in the range of mM. Acquiring a statistically significant result for the above experiment suggests that SAM is accumulating in high concentration inside the cells, which is expected from the reported rate of uptake, but this is not generally the case for products of other enzymatic processes. Therefore, a facile approach with the ability to monitor more subtle differences in product concentrations should be adopted. A solution can be formed with the use of other RS variants such as CARS or SRS. Developing such methods for DE experiments is promising but requires laborious steps which are out of reach for this project's time-line. However, Raman spectral acquisitions of the products of fluorinase, namely 5'-FDA and 5'-ClDA, have not been previously reported and could provide useful insight to the configuration of future CARS or

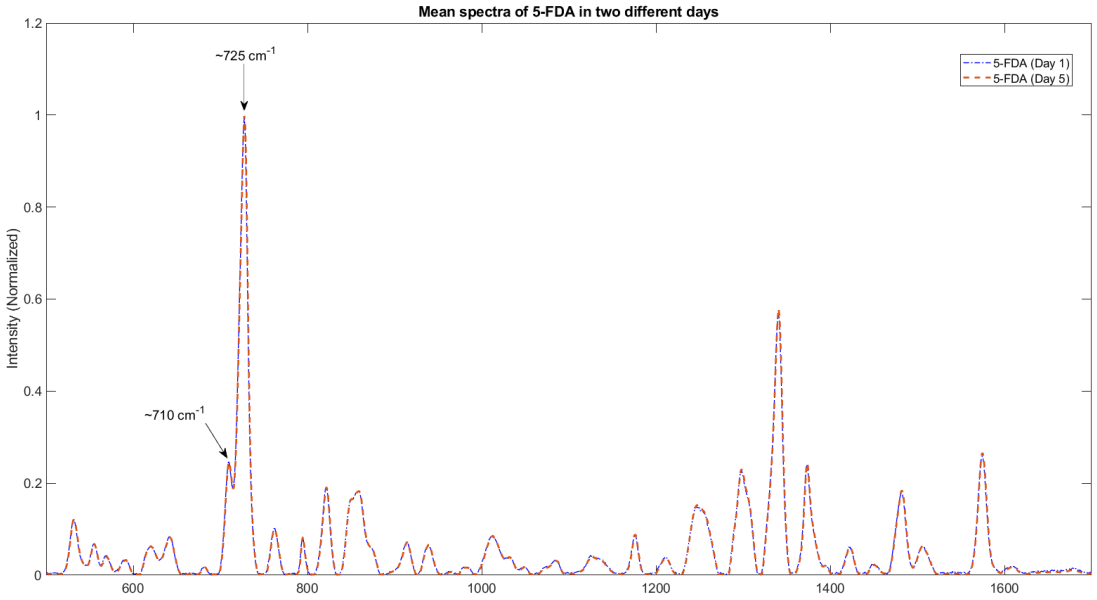


Figure 6.13: Raman Spectrum of 5'-FDA.

SRS experiments. Therefore, Raman spectra of these compounds in powder form, kindly provided by Prof. O'Hagan and Dr Philip Lowe, are presented.

6.3.1 5'-FDA

A small quantity of 5'-FDA powder was placed on a CaF_2 substrate and several Raman spectra were acquired in randomly selected locations. The substrate was left in a close box in aerated conditions (room temperature) and the process was repeated in 5 days. After baseline and cosmic ray spike removal, mean Raman spectra of 5'-FDA in different days are shown in 6.13. Fist of all it is important to note both the stability of the molecule and the robustness of Raman spectral acquisition and spectrum processing technique utilized, as the final spectra virtually coincide between different measurements (mean error is less than 0.5%). The highest peak of 5'-FDA lies in 725 cm^{-1} as it does in SAM. This is expected as it is produced from the adenosine part of the molecule and monitoring this peak will not provide information of the enzymatic transformation

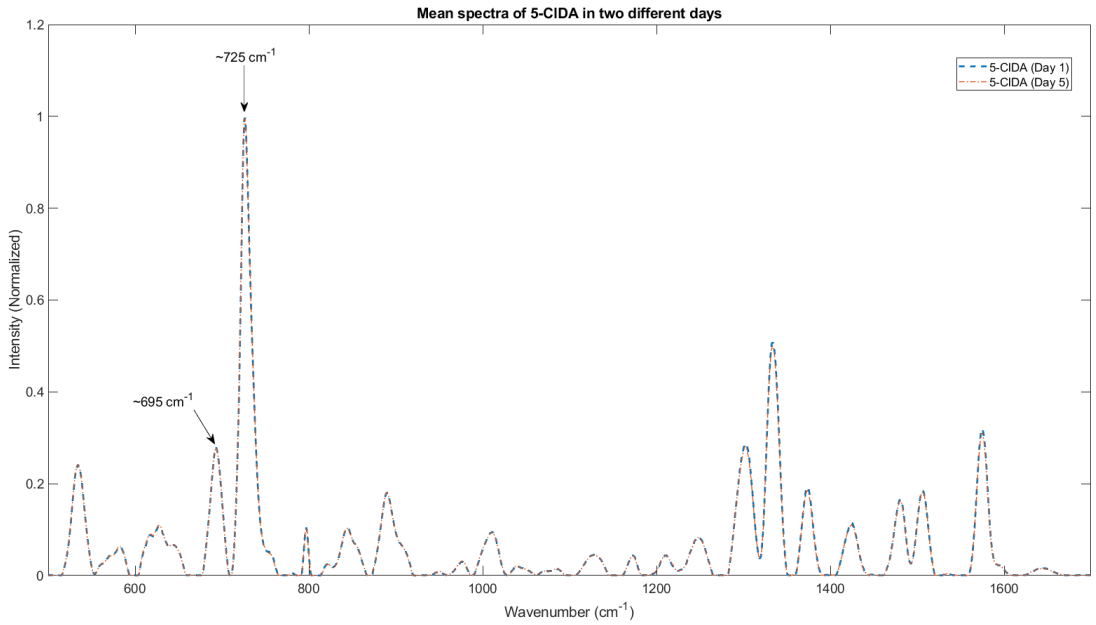


Figure 6.14: Raman Spectrum of 5-ClDA.

of SAM to 5'-FDA. The absence of a peak in 680 cm^{-1} and appearance in 710 cm^{-1} is interesting as it is consistent with the fact that the 680 cm^{-1} peak was produced from the C-S bonds at the interface of L-methionine and adenosine parts of the molecule. From the structure of 5'-FDA, (methionine replaced by fluoride) the peak appearing at 710 cm^{-1} can be tentatively assigned to the C-F bond. Confirmation of this is provided in the next section where comparison with a "brother" molecule is made.

6.3.2 5'-ClDA

Raman spectra of 5'-ClDA were acquired with a similar process as above (figure 6.14). This serves mostly as a comparison to 5'-FDA for distinguishing C-F related peaks. The only difference between the molecules is that the fluoride atom is replaced by a chloride atom. Therefore, any differences between their spectra are attributed to the C-Cl bond substituting C-F. There are many similarities and some differences between spectra, which is to be expected as the molecules are very

similar. Again the adenosine peak (725 cm^{-1}) is dominant in the spectrum. Most of the peaks appear in both spectra with some of them in varying intensities. The only difference with considerable intensity however, is the 710 cm^{-1} peak which appears to have shifted to 695 cm^{-1} and this is a confirmation of the initial assignment of this bond to C-F in the 5'-FDA spectrum. This peak also differs in wavelength with the "nearby" peak seen in SAM by 30 wavenumbers. Therefore, it can serve as a distinguishing attribute in Raman configurations for future experiments aspiring to monitor 5'-FDA intracellular production.

Chapter 7

Concluding remarks and future prospects

In section 2.3.3, four genes with high sequence similarity to the established fluorinases, found by genome mining are reported. Especially the instances encountered in two genomes, *A. mzabensis* and *Amycolatopsis* sp. CA-128772, bear sequence identities comparable to the one among the other fluorinases that have been characterized ($\approx$ 80%) and they are very likely to be fluorinases themselves. This is also strengthened by the fact that there are genes in the genomes of the aforementioned species with identities exceeding 50% in most of the genes that belong to the 4-FT pathway. However, studies of enzyme characterization have not been published for the products of these genes and although current works are probably in progress, fluorinase enzyme characterization is certainly a promising novelty hunting ground.

The analytical approach for modelling growth curves and fluorescence presented in chapter three, harnesses the capacity of *GP* for the regression of experimental data, which has been recently shown to approximate derivatives of noisy biological experimental data more accurately [71]. The translation of a particular FP is

influenced by other elements in the sequence such as the RBS [162] and terminator sequences [163], [164]. Yet, most studies still use the absolute or relative value of fluorescence to show the differences between promoter activities. The proposed metric is not only quantitatively more accurate, but can also act to enable the unravelling of observed differences between alternative RBS and terminators. The improvement in quantification capabilities of regulatory elements can be expanded by constructing a more accurate model, for example by modelling the degradation rate as a function of time rather than a constant based on the availability of the proteases responsible for degradation or other internal processes [165], incorporation of the copy number of the plasmid which hosts the fusion of promoter and FP, or the strength of a particular FP.

The collection of strains that have been constructed (see tables 4.7 and 4.8), used for the experimental validation assays of *in vivo* fluorination and subsequent growth curve analysis that is presented in chapter five, comprise a solid basis for future pathway engineering experiments. The realization of *in vivo* fluorination in *E. coli* by fluorinase function is undoubtedly a major unprecedented result and raises both questions and possibilities. First of all, the positive effect of fluorinase expression in *E. coli* growth shown in chapter five, is an enabling feature for the production of fluorometabolites in bioreactors without suffering from genetic instability issues and is strengthened by the apparent orthogonality of the actual reaction catalysed in strains producing 5'-FDA. The achieved intracellular concentration of the latter (see 5.4.3), provides additional indirect evidence, in addition to the Raman spectral acquisitions, that the constitutively expressed SAM transporter in *E. coli*, is able to uptake SAM in higher concentrations intracellularly, even with increased molarities of extracellularly provided SAM. Enzymatic flurorination is a reversible reaction and, as reported in section 5.4, no exportation of 5'-FDA from the cell is detectable, therefore the concentration of 5'-FDA reaches a maximum corresponding to the chemical equilibrium between the substrates and products. In particular, presence of L-methionine, which

is one of the products, shifts the chemical equilibrium towards the substrates. This represents a limit in the *in vivo* fluorination capability demonstrated by the engineered *E. coli*. In the natural occurrence of the fluorination pathway in *S. cattleya*, there is a flux of the final product (4-FT) extracellularly. The establishment of a similar exportation mechanism of fluorinated products would therefore constitute a desirable attribute for enabling the use of *E. coli* as a microbial factory for fluorometabolites. The implementation can either be in the form of a transporter for 5'-FDA, or as another enzymatic process converting 5'-FDA to an exportable fluorometabolite.

Regarding the development of a screening method, growth curve analysis performed in chapter 5, reveals some synergistic effects of the fluorinase and fluoride, however, this effect was present in strains that did not appear to produce 5'-FDA. The relation of 5'-FDA production with growth effects is therefore deemed inconclusive. This is not surprising as the chemical equilibrium intracellularly does not permit a continuous flux in order to alleviate fluoride toxicity, leading to a growth advantage. In this manner, establishing a flux with an irreversible reaction or exportation of inert fluorinated products would also be an enabling feature towards genetic selection of improved variants. RS reveals differences in spectra between substrates and products, specifically, the C-F assigned peak. The ability to detect intracellular metabolites has also been demonstrated, specifically for SAM, one of the substrates which accumulates with high molarities in *E. coli* by function of the SAM transporter. Unfortunately, the limit of detection of spontaneous Raman scattering lies in tens of milimolars, concentration levels that are forbidding for detection of the minute differences required. However, alternative Raman-based techniques than provide much lower detection limits can be potentially utilized for monitoring intracellular production of 5'-FDA such as CARS, SRS and Surface Enhanced Raman Spectroscopy (SERS). The latter in particular, has been shown to exhibit sensitivity of less than micro-molar concentrations in the detection of adenosine, a similar molecule to 5'-FDA [159].

This project's main deliverable is an unprecedented engineered *E. coli* strain for *in vivo* direct fluorination, accomplished by a recombinant fluorinase. Additionally, the path towards construction of the engineered *E. coli* for fluorinated production yielded a collection of strains that were useful in addressing questions relevant to the effects of the fluorinase, intracellular SAM capacity and the potential of Raman to monitor key metabolites. The Gaussian fitting methods for growth curve analysis presented in chapter three were extensively utilized in chapter five, however, the analytical approach for promoter activities constitutes a stand-alone original piece of work. In general, the outcome of this project is deemed successful.

References

[1] L. Ma, Y. Li, L. Meng, H. Deng, Y. Li, Q. Zhang, and A. Diao, "Biological fluorination from the sea: discovery of a sam-dependent nucleophilic fluorinating enzyme from the marine-derived bacterium streptomyces xinghaiensis nrrl b24674," *RSC Advances*, vol. 6, no. 32, pp. 27 047–27 051, 2016.

[2] T. Matsuura and T. Yomo, "In vitro evolution of proteins," *Journal of bioscience and bioengineering*, vol. 101, no. 6, pp. 449–456, 2006.

[3] C. D. Cadicamo, J. Courtieu, H. Deng, A. Meddour, and D. O'Hagan, "Enzymatic fluorination in streptomyces cattleya takes place with an inversion of configuration consistent with an sn2 reaction mechanism," *ChemBioChem*, vol. 5, no. 5, pp. 685–690, 2004.

[4] H. Deng, S. L. Cobb, A. R. McEwan, R. P. McGlinchey, J. H. Naismith, D. O'Hagan, D. A. Robinson, and J. B. Spencer, "The fluorinase from streptomyces cattleya is also a chlorinase," *Angewandte Chemie International Edition*, vol. 45, no. 5, pp. 759–762, 2006.

[5] A. S. Eustáquio, D. O'Hagan, and B. S. Moore, "Engineering fluorometabolite production: fluorinase expression in salinispora tropica yields fluorosalinosporamide," *Journal of natural products*, vol. 73, no. 3, pp. 378–382, 2010.

[6] A. Kuzminov, "Homologous recombination—experimental systems, analysis and significance," *EcoSal Plus*, vol. 4, no. 2, 2011.

[7] R. Gautam, "Vibrational microspectroscopic studies of biomedical conditions using model systems," Ph.D. dissertation, INDIAN INSTITUTE OF SCIENCE Bangalore-560012, India, 2014.

[8] K. A. Reid, R. D. Bowden, L. Dasaradhi, M. R. Amin, and D. B. Harper, "Biosynthesis of fluorinated secondary metabolites by streptomyces cattleya," *Microbiology*, vol. 141, no. 6, pp. 1385–1393, 1995.

[9] J. L. Baker, N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge, and R. R. Breaker, "Widespread genetic switches and toxicity resistance proteins for fluoride," *Science*, vol. 335, no. 6065, pp. 233–235, 2012.

[10] D. Tsikritsis, "Vibrational spectroscopy and microscopy in colorectal cancer," 2018.

[11] E. Carbonnel, T. Poisson, P. Jubault, X. Pannecoucke, and T. Besset, "Recent advances for the direct introduction of the cf2me moiety," *Frontiers in chemistry*, vol. 7, 2019.

[12] W. Hong, *Agricultural products based on fluorinated heterocyclic compounds*. John Wiley & Sons: Hoboken, 2009.

[13] J. Wang, M. Sánchez-Roselló, J. L. Aceña, C. del Pozo, A. E. Sorochinsky, S. Fustero, V. A. Soloshonok, and H. Liu, "Fluorine in pharmaceutical industry: fluorine-containing drugs introduced to the market in the last decade (2001–2011)," *Chemical reviews*, vol. 114, no. 4, pp. 2432–2506, 2013.

[14] D. O'Hagan, "Fluorine in health care: Organofluorine containing blockbuster drugs," *Journal of Fluorine Chemistry*, vol. 131, no. 11, pp. 1071–1081, 2010.

[15] Y. Zhou, J. Wang, Z. Gu, S. Wang, W. Zhu, J. L. Aceña, V. A. Soloshonok, K. Izawa, and H. Liu, "Next generation of fluorine-containing pharmaceuticals, compounds currently in phase ii–iii clinical trials of major pharmaceutical companies: new structural trends and therapeutic areas," *Chemical reviews*, vol. 116, no. 2, pp. 422–518, 2016.

[16] D. B. Longley, D. P. Harkin, and P. G. Johnston, "5-fluorouracil: mechanisms of action and clinical strategies," *Nature Reviews Cancer*, vol. 3, no. 5, p. 330, 2003.

[17] L. N. Herrera-Rodriguez, H.-P. MEYER, K. T. Robins, and F. KHAN, "Perspectives on biotechnological halogenation," *chimica oggi/Chemistry Today*, vol. 29, no. 6, 2011.

[18] K. B. Reed and H. S. Alper, "Expanding beyond canonical metabolism: interfacing alternative elements, synthetic biology, and metabolic engineering," *Synthetic and systems biotechnology*, 2017.

[19] D. O'Hagan, "Recent developments on the fluorinase from streptomyces cattleya," *Journal of Fluorine Chemistry*, vol. 127, no. 11, pp. 1479–1483, 2006.

[20] K. Markakis, A. De Las Heras, and A. Elfick, "Analytical approach for the calculation of promoter activities based on fluorescent protein expression data," *Engineering Biology*, vol. 1, no. 2, pp. 77–85, 2017.

[21] K. Göpfrich, I. Platzman, and J. P. Spatz, "Mastering complexity: towards bottom-up construction of multifunctional eukaryotic synthetic cells," *Trends in biotechnology*, 2018.

[22] "Registry of standard biological parts," http://parts.igem.org/, accessed: 2018-11-25.

[23] R. S. Cox, C. Madsen, J. A. McLaughlin, T. Nguyen, N. Roehner, B. Bartley, J. Beal, M. Bissell, K. Choi, K. Clancy *et al.*, "Synthetic biology open language (sbol) version 2.2. 0," *Journal of integrative bioinformatics*, vol. 15, no. 1, 2018.

[24] D. O'Hagan and H. Deng, "Enzymatic fluorination and biotechnological developments of the fluorinase," *Chemical reviews*, vol. 115, no. 2, pp. 634–649, 2014.

[25] J. Marais, "Monofluoroacetic acid, the toxic principle of "gifblaar", dichapetalum cymosum (hook) engl." 1944.

[26] P. Oelrichs and T. McEwan, "Isolation of the toxic principle in acacia georginae," *Nature*, vol. 190, no. 4778, pp. 808–809, 1961.

[27] M. M. de Oliveira, "Chromatographic isolation of monofluoroacetic acid from palicourea marcgravii st. hil," *Experientia*, vol. 19, no. 11, pp. 586–587, 1963.

[28] J. Lovelace, G. Miller, and G. Welkie, "The accumulation of fluoroacetate and fluorocitrate in forage crops collected near a phosphate plant," *Atmospheric Environment (1967)*, vol. 2, no. 2, pp. 187–190, 1968.

[29] J. Y. Cheng, M.-H. Yu, G. W. Miller, and G. W. Welkie, "Fluoroorganic acids in soybean leaves exposed to fluoride," *Environmental Science & Technology*, vol. 2, no. 5, pp. 367–370, 1968.

[30] D. O'Hagan, R. Perry, J. M. Lock, J. M. Meyer, L. Dasaradhi, J. T. Hamilton, and D. B. Harper, "High levels of monofluoroacetate in dichapetalum braunii," *Phytochemistry*, vol. 33, no. 5, pp. 1043–1045, 1993.

[31] M. F. Carvalho and R. S. Oliveira, "Natural production of fluorinated compounds and biotechnological prospects of the fluorinase enzyme," *Critical reviews in biotechnology*, vol. 37, no. 7, pp. 880–897, 2017.

[32] C. D. Murphy, C. Schaffrath, and D. O'Hagan, "Fluorinated natural products: the biosynthesis of fluoroacetate and 4-fluorothreonine in streptomyces cattleya," *Chemosphere*, vol. 52, no. 2, pp. 455–461, 2003.

[33] M. C. Walker and M. C. Chang, "Natural and engineered biosynthesis of fluorinated natural products," *Chemical Society Reviews*, vol. 43, no. 18, pp. 6527–6536, 2014.

[34] S. Thomas, V. Singleton, J. Lowery, R. Sharpe, L. Pruess, J. Porter, J. Mowat, N. Bohonos *et al.*, "Nucleocidin, a new antibiotic with activity against trypanosomes." *Nucleocidin, a new antibiotic with activity against trypanosomes.*, 1957.

[35] G. O. Morton, J. E. Lancaster, G. E. Van Lear, W. Fulmor, and W. E. Meyer, "Structure of nucleocidin. iii. revised structure," *Journal of the American Chemical Society*, vol. 91, no. 6, pp. 1535–1537, 1969.

[36] X. M. Zhu, S. Hackl, M. N. Thaker, L. Kalan, C. Weber, D. S. Urgast, E. M. Krupp, A. Brewer, S. Vanner, A. Szawiola *et al.*, "Biosynthesis of the fluorinated natural product nucleocidin in streptomyces calvus is dependent on the blda-specified leu-trnauua molecule," *ChemBioChem*, vol. 16, no. 17, pp. 2498–2506, 2015.

[37] A. Bartholomé, J. E. Janso, D. O'Hagan *et al.*, "Fluorometabolite biosynthesis: isotopically labelled glycerol incorporations into the antibiotic nucleocidin in streptomyces calvus," *Organic & biomolecular chemistry*, vol. 15, no. 1, pp. 61–64, 2017.

[38] M. SANADA, T. MIYANO, S. IWADARE, J. M. WILLIAMSON, B. H. ARISON, J. L. SMITH, A. W. DOUGLAS, J. M. LIESCH, and E. INAMINE, "Biosynthesis of fluorothreonine and fluoroacetic," *The Journal of antibiotics*, vol. 39, no. 2, pp. 259–265, 1986.

[39] H. Deng, S. M. Cross, R. P. McGlinchey, J. T. Hamilton, and D. O'Hagan, "In vitro reconstituted biotransformation of 4-fluorothreonine from fluoride ion: application of the fluorinase," *Chemistry & biology*, vol. 15, no. 12, pp. 1268–1276, 2008.

[40] D. O'hagan, C. Schaffrath, S. L. Cobb, J. T. Hamilton, and C. D. Murphy, "Biochemistry: biosynthesis of an organofluorine molecule," *Nature*, vol. 416, no. 6878, p. 279, 2002.

[41] C. Schaffrath, H. Deng, and D. O'Hagan, "Isolation and characterisation of 5'-fluorodeoxyadenosine synthase, a fluorination enzyme from streptomyces cattleya," *FEBS letters*, vol. 547, no. 1-3, pp. 111–114, 2003.

[42] S. L. Cobb, H. Deng, J. T. Hamilton, R. P. McGlinchey, and D. O'Hagan, "Identification of 5-fluoro-5-deoxy-d-ribose-1-phosphate as an intermediate in fluorometabolite biosynthesis in streptomyces cattleya," *Chemical Communications*, no. 5, pp. 592–593, 2004.

[43] S. L. Cobb, H. Deng, J. T. Hamilton, R. P. McGlinchey, D. O'Hagan, and C. Schaffrath, "The identification of 5'-fluoro-5-deoxyinosine as a shunt product in cell free extracts of streptomyces cattleya," *Bioorganic chemistry*, vol. 33, no. 5, pp. 393–401, 2005.

[44] M. Onega, R. P. McGlinchey, H. Deng, J. T. Hamilton, and D. O'Hagan, "The identification of (3r, 4s)-5-fluoro-5-deoxy-d-ribulose-1-phosphate as an intermediate in fluorometabolite biosynthesis in streptomyces cattleya," *Bioorganic chemistry*, vol. 35, no. 5, pp. 375–385, 2007.

[45] C. D. Murphy, S. J. Moss, and D. O'Hagan, "Isolation of an aldehyde dehydrogenase involved in the oxidation of fluoroacetaldehyde to fluoroacetate instreptomyces cattleya," *Applied and environmental microbiology*, vol. 67, no. 10, pp. 4919–4921, 2001.

[46] K. K. Ho and H. Weiner, "Isolation and characterization of an aldehyde dehydrogenase encoded by the aldb gene of escherichia coli," *Journal of bacteriology*, vol. 187, no. 3, pp. 1067–1073, 2005.

[47] F. Huang, S. F. Haydock, D. Spiteller, T. Mironenko, T.-L. Li, D. O'Hagan, P. F. Leadlay, and J. B. Spencer, "The gene cluster for fluorometabolite biosynthesis in streptomyces cattleya: a thioesterase confers resistance to fluoroacetyl-coenzyme a," *Chemistry & biology*, vol. 13, no. 5, pp. 475–484, 2006.

[48] J. L. McMurry and M. C. Chang, "Fluorothreonyl-trna deacylase prevents mistranslation in the organofluorine producer streptomyces cattleya," *Proceedings of the National Academy of Sciences*, vol. 114, no. 45, pp. 11 920–11 925, 2017.

[49] C. Dong, H. Deng, M. Dorward, C. Schaffrath, D. O'Hagan, and J. H. Naismith, "Crystallization and x-ray diffraction of 5'-fluoro-5'-deoxyadenosine synthase, a fluorination enzyme from streptomyces cattleya," *Acta Crystallographica Section D: Biological Crystallography*, vol. 59, no. 12, pp. 2292–2293, 2003.

[50] C. Dong, F. Huang, H. Deng, C. Schaffrath, J. B. Spencer, D. O'hagan, and J. H. Naismith, "Crystal structure and mechanism of a bacterial fluorinating enzyme," *Nature*, vol. 427, no. 6974, p. 561, 2004.

[51] N. Halliday, K. Hardie, P. Williams, K. Winzer, and D. Barrett, "Quantitative liquid chromatography–tandem mass spectrometry profiling of activated methyl cycle metabolites involved in luxs-dependent quorum sensing in escherichia coli," *Analytical biochemistry*, vol. 403, no. 1-2, pp. 20–29, 2010.

[52] H. Sun, W. L. Yeo, Y. H. Lim, X. Chew, D. J. Smith, B. Xue, K. P. Chan, R. C. Robinson, E. G. Robins, H. Zhao *et al.*, "Directed evolution of a fluorinase for improved fluorination efficiency with a non-native substrate," *Angewandte Chemie*, vol. 128, no. 46, pp. 14 489–14 492, 2016.

[53] H. Deng, L. Ma, N. Bandaranayaka, Z. Qin, G. Mann, K. Kyeremeh, Y. Yu, T. Shepherd, J. H. Naismith, and D. O'Hagan, "Identification of fluorinases from streptomyces sp ma37, norcardia brasiliensis, and actinoplanes sp n902-109 by genome mining," *ChemBioChem*, vol. 15, no. 3, pp. 364–368, 2014.

[54] L. Ma, A. Bartholome, M. H. Tong, Z. Qin, Y. Yu, T. Shepherd, K. Kyeremeh, H. Deng, and D. O'Hagan, "Identification of a fluorometabolite from streptomyces sp. ma37:(2 r 3 s 4 s)-5-fluoro-2, 3, 4-trihydroxypentanoic acid," *Chemical science*, vol. 6, no. 2, pp. 1414–1419, 2015.

[55] M. HimáTong *et al.*, "Fluoroacetate biosynthesis from the marine-derived bacterium streptomyces xinghaiensis nrrl b-24674," *Organic & biomolecular chemistry*, vol. 12, no. 27, pp. 4828–4831, 2014.

[56] B. W. Thuronyi, T. M. Privalsky, and M. C. Chang, "Engineered fluorine metabolism and fluoropolymer production in living cells," *Angewandte Chemie International Edition*, vol. 56, no. 44, pp. 13 637–13 640, 2017.

[57] J. Courcelle, B. M. Wendel, D. D. Livingstone, and C. T. Courcelle, "Recbcd is required to complete chromosomal replication: Implications for double-strand break frequencies and repair mechanisms," *DNA repair*, vol. 32, pp. 86–95, 2015.

[58] C. M. Hamilton, M. Aldea, B. Washburn, P. Babitzke, and S. Kushner, "New method for generating deletions and gene replacements in escherichia coli." *Journal of bacteriology*, vol. 171, no. 9, pp. 4617–4622, 1989.

[59] H. M. Ellis, D. Yu, T. DiTizio *et al.*, "High efficiency mutagenesis, repair, and engineering of chromosomal dna using single-stranded oligonucleotides," *Proceedings of the National Academy of Sciences*, vol. 98, no. 12, pp. 6742–6746, 2001.

[60] R. Benzinger, L. Enquist, and A. Skalka, "Transfection of escherichia coli spheroplasts. v. activity of recbc nuclease in rec+ and rec minus spheroplasts measured with different forms of bacteriophage dna." *Journal of virology*, vol. 15, no. 4, pp. 861–871, 1975.

[61] S. K. Sharan, L. C. Thomason, S. G. Kuznetsov, D. L. Court *et al.*, "Recombineering: a homologous recombination-based method of genetic engineering," *Nature protocols*, vol. 4, no. 2, p. 206, 2009.

[62] H. Bimboim and J. Doly, "A rapid alkaline extraction procedure for screening recombinant plasmid dna," *Nucleic acids research*, vol. 7, no. 6, pp. 1513–1523, 1979.

[63] A. Asif, H. Mohsin, R. Tanvir, and Y. Rehman, "Revisiting the mechanisms involved in calcium chloride induced bacterial transformation," *Frontiers in microbiology*, vol. 8, p. 2169, 2017.

[64] R. Saiki, U. Gyllenstein, and H. Erlich, "Polymerase chain reaction," *Science*, vol. 239, no. 4839, pp. 487–491, 1988.

[65] D. I. Ellis, D. P. Cowcher, L. Ashton, S. O'Hagan, and R. Goodacre, "Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool," *Analyst*, vol. 138, no. 14, pp. 3871–3884, 2013.

[66] W. E. Huang, M. Li, R. M. Jarvis, R. Goodacre, and S. A. Banwart, "Shining light on the microbial world: the application of raman microspectroscopy," in *Advances in applied microbiology*. Elsevier, 2010, vol. 70, pp. 153–186.

[67] X. Zhang, M. B. Roeffaers, S. Basu, J. R. Daniele, D. Fu, C. W. Freudiger, G. R. Holtom, and X. S. Xie, "Label-free live-cell imaging of nucleic acids using stimulated raman scattering microscopy," *ChemPhysChem*, vol. 13, no. 4, pp. 1054–1059, 2012.

[68] G. I. Petrov, R. Arora, V. V. Yakovlev, X. Wang, A. V. Sokolov, and M. O. Scully, "Comparison of coherent and spontaneous raman microspectroscopies for noninvasive detection of single bacterial endospores," *Proceedings of the National Academy of Sciences*, vol. 104, no. 19, pp. 7776–7779, 2007.

[69] J. Monod, "The growth of bacterial cultures," *Annual Reviews in Microbiology*, vol. 3, no. 1, pp. 371–394, 1949.

[70] M. Zwietering, I. Jongenburger, F. Rombouts, and K. Van't Riet, "Modeling of the bacterial growth curve," *Applied and environmental microbiology*, vol. 56, no. 6, pp. 1875–1881, 1990.

[71] P. S. Swain, K. Stevenson, A. Leary, L. F. Montano-Gutierrez, I. B. Clark, J. Vogel, and T. Pilizota, "Inferring time derivatives including cell growth rates using gaussian processes," *Nature communications*, vol. 7, p. 13766, 2016.

[72] A. De Las Heras, W. Xiao, V. Sren, and A. Elfick, "Edwin: A robotic platform for automated rna extraction and analysis during reporter gene–based dynamic characterization of bacterial promoters," *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, vol. 22, no. 1, pp. 50–62, 2017.

[73] D. L. Shis and M. R. Bennett, "Library of synthetic transcriptional and gates built with split t7 rna polymerase mutants," *Proceedings of the National Academy of Sciences*, vol. 110, no. 13, pp. 5028–5033, 2013.

[74] T. J. Rudge, J. R. Brown, F. Federici, N. Dalchau, A. Phillips, J. W. Ajioka, and J. Haseloff, "Characterization of intrinsic properties of promoters," *ACS synthetic biology*, vol. 5, no. 1, pp. 89–98, 2016.

[75] H. Schwetlick and T. Schütze, "Least squares approximation by splines with free knots," *BIT Numerical mathematics*, vol. 35, no. 3, pp. 361–384, 1995.

[76] T. Schütze and H. Schwetlick, "Constrained approximation by splines with free knots," *BIT Numerical Mathematics*, vol. 37, no. 1, pp. 105–137, 1997.

[77] X. Wang, B. Errede, and T. C. Elston, "Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters," *Biophysical journal*, vol. 94, no. 6, pp. 2017–2026, 2008.

[78] A. Bren, Y. Hart, E. Dekel, D. Koster, and U. Alon, "The last generation of bacterial growth in limiting nutrient," *BMC systems biology*, vol. 7, no. 1, p. 27, 2013.

[79] M. Ronen, R. Rosenberg, B. I. Shraiman, and U. Alon, "Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics," *Proceedings of the national academy of sciences*, vol. 99, no. 16, pp. 10 555–10 560, 2002.

[80] L. Keren, O. Zackay, M. Lotan-Pompan, U. Barenholz, E. Dekel, V. Sasson, G. Aidelberg, A. Bren, D. Zeevi, A. Weinberger *et al.*, "Promoters maintain their relative activity levels under different growth conditions," *Molecular systems biology*, vol. 9, no. 1, p. 701, 2013.

[81] C. K. Williams and C. E. Rasmussen, "Gaussian processes for machine learning," *the MIT Press*, vol. 2, no. 3, p. 4, 2006.

[82] J. Sambrook, E. F. Fritsch, T. Maniatis *et al.*, *Molecular cloning: a laboratory manual*. Cold spring harbor laboratory press, 1989, no. Ed. 2.

[83] R. S. Cox, M. J. Dunlop, and M. B. Elowitz, "A synthetic three-color scaffold for monitoring genetic regulation and noise," *Journal of biological engineering*, vol. 4, no. 1, p. 10, 2010.

[84] R. Silva-Rocha, E. Martínez-García, B. Calles, M. Chavarría, A. Arce-Rodríguez, A. de las Heras, A. D. Paez-Espino, G. Durante-Rodríguez, J. Kim, P. I. Nikel *et al.*, "The standard european vector architecture (seva): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes," *Nucleic acids research*, vol. 41, no. D1, pp. D666–D675, 2012.

[85] "Anderson promoter collection," http://parts.igem.org/Promoters/Catalog/Anderson, accessed: 2018-10-09.

[86] G. R. Amores, A. de las Heras, A. Sanches-Medeiros, A. Elfick, and R. Silva-Rocha, "Systematic identification of novel regulatory interactions controlling biofilm formation in the bacterium escherichia coli," *Scientific reports*, vol. 7, no. 1, p. 16768, 2017.

[87] J. Chappell, K. Jensen, and P. S. Freemont, "Validation of an entirely in vitro approach for rapid prototyping of dna regulatory elements for synthetic biology," *Nucleic acids research*, vol. 41, no. 5, pp. 3471–3481, 2013.

[88] M. Mauri and S. Klumpp, "A model for sigma factor competition in bacterial cells," *PLoS computational biology*, vol. 10, no. 10, p. e1003845, 2014.

[89] M. Jishage and A. Ishihama, "A stationary phase protein in escherichia coli with binding activity to the major σ subunit of rna polymerase," *Proceedings of the National Academy of Sciences*, vol. 95, no. 9, pp. 4953–4958, 1998.

[90] V. de Lorenzo, L. Eltis, B. Kessler, and K. N. Timmis, "Analysis of pseudomonas gene products using laciq/ptrp-lac plasmids and transposons that confer conditional phenotypes," *Gene*, vol. 123, no. 1, pp. 17–24, 1993.

[91] H. C. Winther-Larsen, K. D. Josefsen, T. Brautaset, and S. Valla, "Parameters affecting gene expression from the pm promoter in gram-negative bacteria," *Metabolic engineering*, vol. 2, no. 2, pp. 79–91, 2000.

[92] A. Marbach and K. Bettenbrock, "lac operon induction in escherichia coli: Systematic comparison of iptg and tmg induction and influence of the transacetylase laca," *Journal of biotechnology*, vol. 157, no. 1, pp. 82–88, 2012.

[93] R. Y. Tsien, "The green fluorescent protein," 1998.

[94] E. L. Snapp, "Fluorescent proteins: a cell biologist's user guide," *Trends in cell biology*, vol. 19, no. 11, pp. 649–655, 2009.

[95] J. T. ANDERSEN, P. POULSEN, and K. F. JENSEN, "Attenuation in the rph-pyre operon of escherichia coli and processing of the dicistronic mrna," *European journal of biochemistry*, vol. 206, no. 2, pp. 381–390, 1992.

[96] J. B. Andersen, C. Sternberg, L. K. Poulsen, S. P. Bjørn, M. Givskov, and S. Molin, "New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria," *Applied and environmental microbiology*, vol. 64, no. 6, pp. 2240–2246, 1998.

[97] J. R. Houser, E. Ford, S. M. Chatterjea, S. Maleri, T. C. Elston, and B. Errede, "An improved short-lived fluorescent protein transcriptional reporter for saccharomyces cerevisiae," *Yeast*, vol. 29, no. 12, pp. 519–530, 2012.

[98] F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez, *Mathematical models in population biology and epidemiology*. Springer, 2012, vol. 40.

[99] A.-C. Freydank, W. Brandt, and B. Dräger, "Protein structure modeling indicates hexahistidine-tag interference with enzyme activity," *Proteins: Structure, function, and bioinformatics*, vol. 72, no. 1, pp. 173–183, 2008.

[100] Y. Liao, B. W. Brandt, J. Li, W. Crielaard, C. Van Loveren, and D. M. Deng, "Fluoride resistance in streptococcus mutans: a mini review," *Journal of oral microbiology*, vol. 9, no. 1, p. 1344509, 2017.

[101] G. Andreadis, V. Topitsoglou, and S. Kalfas, "Acidogenicity and acidurance of dental plaque and saliva sediment from adults in relation to caries activity and chlorhexidine exposure," *Journal of oral microbiology*, vol. 7, no. 1, p. 26197, 2015.

[102] I. Hamilton, "Biochemical effects of fluoride on oral bacteria," *Journal of dental research*, vol. 69, no. 2_suppl, pp. 660–667, 1990.

[103] J. Qin, G. Chai, J. M. Brewer, L. L. Lovelace, and L. Lebioda, "Fluoride inhibition of enolase: crystal structure and thermodynamics," *Biochemistry*, vol. 45, no. 3, pp. 793–800, 2006.

[104] R. B. Stockbridge, H.-H. Lim, R. Otten, C. Williams, T. Shane, Z. Weinberg, and C. Miller, "Fluoride resistance and transport by riboswitch-controlled clc antiporters," *Proceedings of the National Academy of Sciences*, 2012.

[105] S. Li, K. D. Smith, J. H. Davis, P. B. Gordon, R. R. Breaker, and S. A. Strobel, "Eukaryotic resistance to fluoride toxicity mediated by a widespread family of fluoride export proteins," *Proceedings of the National Academy of Sciences*, vol. 110, no. 47, pp. 19 018–19 023, 2013.

[106] K. H. Hu, E. Liu, K. Dean, M. Gingas, W. DeGraff, and N. J. Trun, "Overproduction of three genes leads to camphor resistance and chromosome condensation in escherichia coli," *Genetics*, vol. 143, no. 4, pp. 1521–1532, 1996.

[107] O. Sand, M. Gingras, N. Beck, C. Hall, and N. Trun, "Phenotypic characterization of overexpression or deletion of the escherichia colicrca, cspe and crcb genes," *Microbiology*, vol. 149, no. 8, pp. 2107–2117, 2003.

[108] R. B. Stockbridge, L. Kolmakova-Partensky, T. Shane, A. Koide, S. Koide, C. Miller, and S. Newstead, "Crystal structures of a double-barrelled fluoride ion channel," *Nature*, vol. 525, no. 7570, p. 548, 2015.

[109] M. Rapp, E. Granseth, S. Seppälä, and G. Von Heijne, "Identification and evolution of dual-topology membrane proteins," *Nature structural and molecular biology*, vol. 13, no. 2, p. 112, 2006.

[110] T. Berbasova, S. Nallur, T. Sells, K. D. Smith, P. B. Gordon, S. L. Tausta, and S. A. Strobel, "Fluoride export (fex) proteins from fungi, plants and animals are'single barreled'channels containing one functional and one vestigial ion pore," *PloS one*, vol. 12, no. 5, p. e0177096, 2017.

[111] R. B. Stockbridge, J. L. Robertson, L. Kolmakova-Partensky, and C. Miller, "A family of fluoride-specific ion channels with dual-topology architecture," *Elife*, vol. 2, p. e01084, 2013.

[112] H. Deng, D. O'Hagan, and C. Schaffrath, "Fluorometabolite biosynthesis and the fluorinase from streptomyces cattleya," *Natural product reports*, vol. 21, no. 6, pp. 773–784, 2004.

[113] C. Ji, R. B. Stockbridge, and C. Miller, "Bacterial fluoride resistance, fluc channels, and the weak acid accumulation effect," *The Journal of general physiology*, vol. 144, no. 3, pp. 257–261, 2014.

[114] G. Pósfai, V. Kolisnychenko, Z. Bereczki, and F. R. Blattner, "Markerless gene replacement in escherichia coli stimulated by a double-strand break in the chromosome," *Nucleic acids research*, vol. 27, no. 22, pp. 4409–4415, 1999.

[115] E. Martínez-García and V. de Lorenzo, "Engineering multiple genomic deletions in gram-negative bacteria: analysis of the multi-resistant antibiotic profile of pseudomonas putida kt2440," *Environmental microbiology*, vol. 13, no. 10, pp. 2702–2716, 2011.

[116] T. Aparicio, V. de Lorenzo, and E. Martínez-García, "Broadening the seva plasmid repertoire to facilitate genomic editing of gram-negative bacteria," in *Hydrocarbon and Lipid Microbiology Protocols*. Springer, 2015, pp. 9–27.

[117] W. B. Parker and E. J. Sorscher, "Use of e. coli purine nucleoside phosphorylase in the treatment of solid tumors," *Current pharmaceutical design*, vol. 23, no. 45, pp. 7003–7024, 2017.

[118] E. M. Bennett, C. Li, P. W. Allan, W. B. Parker, and S. E. Ealick, "Structural basis for substrate specificity of escherichia coli purine nucleoside phosphorylase," *Journal of Biological Chemistry*, vol. 278, no. 47, pp. 47 110–47 118, 2003.

[119] G. Dandanell, R. H. Szczepanowski, B. Kierdaszuk, D. Shugar, and M. Bochtler, "Escherichia coli purine nucleoside phosphorylase ii, the product of the xapa gene," *Journal of molecular biology*, vol. 348, no. 1, pp. 113–125, 2005.

[120] G. W. Koszalka, J. Vanhooke, S. Short, and W. W. Hall, "Purification and properties of inosine-guanosine phosphorylase from escherichia coli k-12." *Journal of bacteriology*, vol. 170, no. 8, pp. 3493–3498, 1988.

[121] R. Buxton, "Genetic analysis of thymidine-resistant and low-thymine-requiring mutants of escherichia coli k-12 induced by bacteriophage mu-1." *Journal of bacteriology*, vol. 121, no. 2, pp. 475–484, 1975.

[122] K. Watanabe, S. Tomioka, K. Tanimura, H. Oku, and K. Isoi, "Uptake of amp, adp, and atp in escherichia coli w," *Bioscience, biotechnology, and biochemistry*, vol. 75, no. 1, pp. 7–12, 2011.

[123] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz, "Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli," *Nature chemical biology*, vol. 5, no. 8, p. 593, 2009.

[124] A. M. Tucker, H. H. Winkler, L. O. Driskell, and D. O. Wood, "S-adenosylmethionine transport in rickettsia prowazekii," *Journal of bacteriology*, vol. 185, no. 10, pp. 3031–3035, 2003.

[125] G. D. Markham and M. A. Pajares, "Structure-function relationships in methionine adenosyltransferases," *Cellular and molecular life sciences*, vol. 66, no. 4, p. 636, 2009.

[126] C. Satishchandran, J. C. Taylor, and G. D. Markham, "Isozymes of s-adenosylmethionine synthetase are encoded by tandemly duplicated genes in escherichia coli," *Molecular microbiology*, vol. 9, no. 4, pp. 835–846, 1993.

[127] Y. Wei and E. Newman, "Studies on the role of the metk gene product of escherichia coli k-12," *Molecular microbiology*, vol. 43, no. 6, pp. 1651–1656, 2002.

[128] E. Newman, L. Budman, E. Chan, R. Greene, R. Lin, C. Woldringh, and R. D'Ari, "Lack of s-adenosylmethionine results in a cell division defect in escherichia coli," *Journal of bacteriology*, vol. 180, no. 14, pp. 3614–3619, 1998.

[129] S. Wang, S. R. Arends, D. S. Weiss, and E. B. Newman, "A deficiency in s-adenosylmethionine synthetase interrupts assembly of the septal ring in escherichia coli k-12," *Molecular microbiology*, vol. 58, no. 3, pp. 791–799, 2005.

[130] L. M. Posnick and L. D. Samson, "Influence of s-adenosylmethionine pool size on spontaneous mutation, dam methylation, and cell growth ofescherichia coli," *Journal of bacteriology*, vol. 181, no. 21, pp. 6756–6762, 1999.

[131] G. Macintyre, C. V. Atwood, and C. G. Cupples, "Lowering s-adenosylmethionine levels inescherichia coli modulates c-to-t transition mutations," *Journal of bacteriology*, vol. 183, no. 3, pp. 921–927, 2001.

[132] A. S. Eustáquio, F. Pojer, J. P. Noel, and B. S. Moore, "Discovery and characterization of a marine bacterial sam-dependent chlorinase," *Nature chemical biology*, vol. 4, no. 1, p. 69, 2008.

[133] H. Deng, S. A. McMahon, A. S. Eustáquio, B. S. Moore, J. H. Naismith, and D. O'Hagan, "Mechanistic insights into water activation in sam hydroxide adenosyltransferase (duf-62)," *ChemBioChem*, vol. 10, no. 15, pp. 2455–2459, 2009.

[134] A. S. Eustáquio, J. Härle, J. P. Noel, and B. S. Moore, "S-adenosyl-l-methionine hydrolase (adenosine-forming), a conserved bacterial and archeal protein related to sam-dependent halogenases," *ChemBioChem*, vol. 9, no. 14, pp. 2215–2219, 2008.

[135] A. Rouillon, Y. Surdin-Kerjan, and D. Thomas, "Transport of sulfonium compounds characterization of the s-adenosylmethionine andsmethylmethionine permeases from the yeastsaccharomyces cerevisiae," *Journal of Biological Chemistry*, vol. 274, no. 40, pp. 28 096–28 105, 1999.

[136] C. Marobbio, G. Agrimi, F. Lasorsa, and F. Palmieri, "Identification and functional reconstitution of yeast mitochondrial carrier for s-adenosylmethionine," *The EMBO Journal*, vol. 22, no. 22, pp. 5975–5982, 2003.

[137] G. Agrimi, M. Di Noia, C. Marobbio, G. Fiermonte, F. Lasorsa, and F. Palmieri, "Identification of the human mitochondrial s-adenosylmethionine transporter: bacterial expression, reconstitution, functional characterization and tissue distribution," *Biochemical Journal*, vol. 379, no. 1, pp. 183–190, 2004.

[138] F. Bouvier, N. Linka, J.-C. Isner, J. Mutterer, A. P. Weber, and B. Camara, "Arabidopsis samt1 defines a plastid transporter regulating plastid biogenesis and plant development," *The Plant Cell*, vol. 18, no. 11, pp. 3088–3105, 2006.

[139] R. Binet, R. E. Fernandez, D. J. Fisher, and A. T. Maurelli, "Identification and characterization of the chlamydia trachomatis l2 s-adenosylmethionine transporter," *MBio*, vol. 2, no. 3, pp. e00 051–11, 2011.

[140] S. G. Andersson, A. Zomorodipour, J. O. Andersson, T. Sicheritz-Pontén, U. C. M. Alsmark, R. M. Podowski, A. K. Näslund, A.-S. Eriksson, H. H. Winkler, and C. G. Kurland, "The genome sequence of rickettsia prowazekii and the origin of mitochondria," *Nature*, vol. 396, no. 6707, p. 133, 1998.

[141] J. Wixon, "Reductive evolution in bacteria: Buchnera sp., rickettsia prowazekii and mycobacterium leprae," *International Journal of Genomics*, vol. 2, no. 1, pp. 44–48, 2001.

[142] L. O. Driskell, A. M. Tucker, H. H. Winkler, and D. O. Wood, "Rickettsial metk-encoded methionine adenosyltransferase expression in an escherichia coli metk deletion strain," *Journal of bacteriology*, vol. 187, no. 16, pp. 5719–5722, 2005.

[143] Z. El-Hajj, R. Reyes-Lamothe, and E. Newman, "Cell division, one-carbon metabolism and methionine synthesis in a metk-deficient escherichia coli mutant, and a role for mmum," *Microbiology*, vol. 159, no. 10, pp. 2036–2048, 2013.

[144] D. H. Figurski and D. R. Helinski, "Replication of an origin-containing derivative of plasmid rk2 dependent on a plasmid function provided in trans," *Proceedings of the National Academy of Sciences*, vol. 76, no. 4, pp. 1648–1652, 1979.

[145] W. B. Wood, "Host specificity of dna produced by escherichia coli: bacterial mutations affecting the restriction and modification of dna," *Journal of molecular biology*, vol. 16, no. 1, pp. 118–IN3, 1966.

[146] K. A. Datsenko and B. L. Wanner, "One-step inactivation of chromosomal genes in escherichia coli k-12 using pcr products," *Proceedings of the National Academy of Sciences*, vol. 97, no. 12, pp. 6640–6645, 2000.

[147] L. Dumon-Seignovert, G. Cariot, and L. Vuillard, "The toxicity of recombinant proteins in escherichia coli: a comparison of overexpression in bl21 (de3), c41 (de3), and c43 (de3)," *Protein expression and purification*, vol. 37, no. 1, pp. 203–206, 2004.

[148] B. Miroux and J. E. Walker, "Over-production of proteins in escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels," *Journal of molecular biology*, vol. 260, no. 3, pp. 289–298, 1996.

[149] S. H. Hong, S. J. Park, S. Y. Moon, J. P. Park, and S. Y. Lee, "In silico prediction and validation of the importance of the entner–doudoroff pathway in poly (3-hydroxybutyrate) production by metabolically engineered escherichia coli," *Biotechnology and bioengineering*, vol. 83, no. 7, pp. 854–863, 2003.

[150] K. E. Tyo, P. K. Ajikumar, and G. Stephanopoulos, "Stabilized gene duplication enables long-term selection-free heterologous pathway expression," *Nature biotechnology*, vol. 27, no. 8, p. 760, 2009.

[151] A. Das, S. Biswas, and M. Biswas, "Expression of phi11 gp07 causes filamentation in escherichia coli," *The open microbiology journal*, vol. 12, p. 107, 2018.

[152] A. Morana, P. Stiuso, G. Colonna, M. Lamberti, M. Carteni, and M. De Rosa, "Stabilization of s-adenosyl-l-methionine promoted by trehalose," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1573, no. 2, pp. 105–108, 2002.

[153] W. B. Whitaker, J. A. Jones, R. K. Bennett, J. E. Gonzalez, V. R. Vernacchio, S. M. Collins, M. A. Palmer, S. Schmidt, M. R. Antoniewicz, M. A. Koffas *et al.*, "Engineering the biological conversion of methanol to specialty chemicals in escherichia coli," *Metabolic engineering*, vol. 39, pp. 49–59, 2017.

[154] "Spectral resolution and spectrometers," https://research.cbc.osu.edu/allen.697/wp-content/uploads/2012/05/Resolution-presentation1.ppt, accessed: 2018-11-25.

[155] C. Liu and R. W. Berg, "Determining the spectral resolution of a charge-coupled device (ccd) raman instrument," *Applied spectroscopy*, vol. 66, no. 9, pp. 1034–1043, 2012.

[156] H. G. Schulze, R. B. Foist, K. Okuda, A. Ivanov, and R. F. Turner, "A small-window moving average-based fully automated baseline estimation method for raman spectra," *Applied spectroscopy*, vol. 66, no. 7, pp. 757–764, 2012.

[157] H. G. Schulze and R. F. Turner, "A fast, automated, polynomial-based cosmic ray spike-removal method for the high-throughput processing of raman spectra," *Applied spectroscopy*, vol. 67, no. 4, pp. 457–462, 2013.

[158] "S-(5'-adenosyl)-l-methionine chloride dihydrochloride," https://www.sigmaaldrich.com/catalog/product/sigma/a7007, accessed: 2018-11-25.

[159] S. Xu, B. Man, S. Jiang, J. Wang, J. Wei, S. Xu, H. Liu, S. Gao, H. Liu, Z. Li *et al.*, "Graphene/cu nanoparticle hybrids fabricated by chemical vapor deposition as surface-enhanced raman scattering substrate for label-free detection of adenosine," *ACS applied materials & interfaces*, vol. 7, no. 20, pp. 10 977–10 987, 2015.

[160] Z. Movasaghi, S. Rehman, and I. U. Rehman, "Raman spectroscopy of biological tissues," *Applied Spectroscopy Reviews*, vol. 42, no. 5, pp. 493–541, 2007.

[161] S. Palchaudhuri, S. J. Rehse, K. Hamasha, T. Syed, E. Kurtovic, E. Kurtovic, and J. Stenger, "Raman spectroscopy of xylitol uptake and

metabolism in gram-positive and gram-negative bacteria," *Applied and environmental microbiology*, vol. 77, no. 1, pp. 131–137, 2011.

[162] B. Reeve, T. Hargest, C. Gilbert, and T. Ellis, "Predicting translation initiation rates for designing synthetic biology," *Frontiers in bioengineering and biotechnology*, vol. 2, p. 1, 2014.

[163] G. Cambray, J. C. Guimaraes, V. K. Mutalik, C. Lam, Q.-A. Mai, T. Thimmaiah, J. M. Carothers, A. P. Arkin, and D. Endy, "Measurement and modeling of intrinsic transcription terminators," *Nucleic acids research*, vol. 41, no. 9, pp. 5139–5148, 2013.

[164] M. Yamanishi, Y. Ito, R. Kintaka, C. Imamura, S. Katahira, A. Ikeuchi, H. Moriya, and T. Matsuyama, "A genome-wide activity assessment of terminator regions in saccharomyces cerevisiae provides a "terminatome" toolbox," *ACS synthetic biology*, vol. 2, no. 6, pp. 337–347, 2013.

[165] I. Levchenko, M. Seidel, R. T. Sauer, and T. A. Baker, "A specificity-enhancing factor for the clpxp degradation machine," *Science*, vol. 289, no. 5488, pp. 2354–2356, 2000.